

Sheffield Hallam University

Artificial Intelligence Techniques and Developments for Infrared Imaging Based Bone Fracture Screening

SHOBAYO, Olamilekan Saheed

Available from the Sheffield Hallam University Research Archive (SHURA) at:

<https://shura.shu.ac.uk/37444/>

A Sheffield Hallam University thesis

This thesis is protected by copyright which belongs to the author.

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Please visit <https://shura.shu.ac.uk/37444/> and <http://shura.shu.ac.uk/information.html> for further details about copyright and re-use permissions.

Artificial Intelligence Techniques and Developments for Infrared Imaging Based Bone Fracture Screening

Olamilekan Saheed Shobayo

A thesis submitted in partial fulfilment of the requirements of
Sheffield Hallam University
for the degree of Doctor of Philosophy

December 2025

Declaration

I hereby declare that:

1. I have not been enrolled for another award of the University, or other academic or professional organisation, whilst undertaking my research degree.
2. None of the material contained in the thesis has been used in any other submission for an academic award.
3. I certify that this thesis is my own work. The use of all published or other sources of material consulted have been properly and fully acknowledged. I confirm that I have sought and obtained copyright permission for any third-party materials included in this thesis.

I used AI at AITS 2 (AI for Shaping) of the Artificial Intelligence Transparency Scale (AITS). I acknowledge the use of Microsoft Copilot <https://copilot.microsoft.com/> to check the spelling and grammar in all sections of my thesis.

4. The work undertaken towards the thesis has been conducted in accordance with the SHU Principles of Integrity in Research and the SHU Research Ethics Policy, and ethics approval has been granted for all research studies in the thesis, as shown in the table below.

Ethics review reference number	Title of research study	Approval date	Date of any post-approval amendments (if applicable)
19/YH/0001	High Resolution Infrared Thermography as a diagnostic aid in paediatric wrist injuries.	07/03/2019	-
ER29956737	Artificial Intelligence Technique Development for Infrared thermal Imaging based bone fracture screening.	09/11/2021	-

5. The word count of the thesis is 53,181.

Name	Olamilekan Saheed Shobayo
Date	December 2025
Award	PhD
College	Business, Technology and Engineering
Director(s) of Studies	Professor Reza Saatchi

Abstract

This study developed and evaluated infrared thermal (IRT) imaging, coupled with artificial intelligence, as a technology for paediatric wrist-injury assessment. It addressed two persistent issues in emergency care: avoidable radiographs for soft-tissue injuries and unnecessary radiation exposure and associated cost. It was established that fractures and sprains which are clinically overlapping at presentation have distinct physiological signatures that manifest as significant temperature differences at the skin surface, which IRT imaging can detect reliably under a controlled acquisition. By developing and testing end-to-end pipelines, this study demonstrated the practical feasibility of IRT imaging-based screening as an adjunct to the current X-ray pathway. A review of conventional imaging (radiography, MRI, ultrasound) was undertaken and then IRT imaging positioned as a technology that can complement the existing diagnostics while reducing unnecessary radiation. Prior studies and early AI applications provided evidence that thermal features support fracture identification that was the focus of the study.

To acquire the IRT images, a rigorous data-collection protocol was designed within Sheffield Children's Hospital, ensuring ethical compliance, recording environmental control, and reproducible camera geometry. Forty children comprising 24 males and 16 females, mean age 10.50 years (standard deviation 2.63 years), 19 with a wrist fracture and 21 with a sprain were included for study with their injury type diagnosis (fracture or sprain) confirmed by an X-ray radiograph. Their mean body temperature was 36.3 °C (standard deviation 0.43 °C) across all participants. Bilateral, 10-second IRT videos were recorded and stabilised; anatomically standardised regions of interest (ROIs) covering carpal bones and distal radius/ulna enabled subject-wise, contralateral comparison that mitigates inter-patient variability. Three complementary feature extraction techniques were used, which included image pixel intensity statistics, a grid-based "hot-spot" representation, and frequency-domain (fast Fourier transform, FFT) magnitude features. These were used as part of training representative models: multilayer perceptron (MLP), convolutional neural network (CNN), adaptive neurofuzzy inference system (ANFIS), with appropriate normalisation, cross-validation, and targeted augmentation.

The study's findings provided a pilot analysis relating temperature to time-since-injury (TSI) and showed no early correlation for fractures but a subsequent moderate relationship, consistent with evolving inflammatory perfusion; sprains exhibited no significant temperature-TSI correlation. Secondly, an MLP trained on grid-based statistics achieved sensitivity 84.2%, specificity 71.4%, and accuracy 77.5%. This is a credible screening performance given the small cohort. Thirdly, a CNN processing FFT-transformed wrist ROIs, supported by rotation/translation/shear augmentation, provided an area under receiver operating characteristic (ROC) curve of 0.82 (accuracy 76%, sensitivity 88%, specificity 68%) and finally, a complementary ANFIS framework, trained on a compact triad of discriminative statistics (standard deviation, interquartile range, kurtosis), achieved effective separation of fractures and sprains while preserving interpretability via Takagi-Sugeno rules and membership functions. The study revealed that K-means-initialised Gaussian memberships offered the best balance of convergence speed and generalisation, outperforming fuzzy-C-means initialisations. Random initialisations of the ANFIS framework provided the best inference of the artificial intelligence models compared.

In conclusion, the carefully standardised IRT, paired with thoughtfully chosen features and models, distinguished paediatric wrist fractures from sprains with clinically relevant performance, biological plausibility, and a credible path to explainable deployment. The evidence supported IRT imaging-based screening as a viable component in a modern fracture-care pathway, promising earlier decision-making, fewer unnecessary radiographs, and improved experience for paediatric care.

Acknowledgement

My profound gratitude goes to the Almighty God, the author and finisher of all good things and for His grace, protection, love, and mercy throughout this doctoral journey. Without His guidance and sustaining hand, this work would not have been possible.

I owe an enduring debt of thanks to my Director of Studies, Professor Reza Saatchi, whose expertise, unwavering support, and exacting yet encouraging supervision shaped this thesis from its earliest ideas to its final form. Your insightful comments, timely suggestions, and commitment to high standards sharpened my thinking and made me a more rigorous and reflective researcher.

My sincere appreciation also goes to my co-supervisor, Professor Shammi Ramlakhan, for his steady guidance, constructive advice, and kindness at every stage. I am equally grateful to Dr. Charlotte Reed for her helpful contribution with the data collection protocols.

I extend heartfelt thanks to all the volunteers who participated in the data-collection sessions, and to the clinical collaborators and support staff who facilitated access, scheduling, and ethical compliance. Your cooperation and goodwill made the research both feasible and rewarding.

To my academic community which includes friends, colleagues, and fellow doctoral researchers, in the School of Computing and MERI, I want to thank you for you for the thoughtful conversations, and the many quiet acts of encouragement that sustained momentum during demanding periods.

Finally, my deepest gratitude goes to all members of my family for their inestimable contribution to my study, especially to Alhaja Sherifat Oyewole, Mrs Remilekun Olawumi, Mr Adekunle Adebisi, and Mr Akeem Atoyebi and every other names not mentioned.

May God richly bless you all.

Dedication

To the cherished memory of my late parents, Alhaja Ramotalahi Omoladun Adebisi and Pa Taiwo Hassan Shobayo. Your boundless love, prayers, and example remain my compass. Though you are not here to witness this milestone, I feel your presence in every step. This work is dedicated to you.

To my wife, Mrs. Adeshola Shobayo thank you for your unwavering support, patience, and faith through long days and longer nights. Your belief in me sustained this journey.

To my children, Samantha, Star, and Saviour Shobayo your laughter kept my purpose clear and my heart full. May this achievement remind you that perseverance, grace, and faith can carry any dream to completion.

With love and gratitude, this thesis is for you.

Publications

Journals

1. **Shobayo, O.**, Saatchi, R., & Ramlakhan, S. (2025). Adaptive Neuro-Fuzzy Inference System Framework for Paediatric Wrist Injury Classification. *Multimodal Technologies and Interaction*, 9(10), 104.
2. **Shobayo, O.**, & Saatchi, R. (2025). Developments in Deep Learning Artificial Neural Network Techniques for Medical Image Analysis and Interpretation. *Diagnostics*, 15(9), 1072.
3. **Shobayo, O.**, Saatchi, R., & Ramlakhan, S. (2024). Convolutional neural network to classify infrared thermal images of fractured wrists in paediatrics. In *Healthcare* (Vol. 12, No. 10, p. 994). MDPI.
4. Tanimola, O., **Shobayo, O.**, Popoola, O., & Okoyeigbo, O. (2024). Breast cancer classification using Fine-Tuned SWIN Transformer model on mammographic images. *Analytics*, 3(4), 461-475.
5. Ogunleye, B., Sharma, H., & **Shobayo, O.** (2024). Sentiment Informed Sentence BERT-Ensemble Algorithm for Depression Detection. *Big Data and Cognitive Computing*, 8(9), 112.
6. **Shobayo, O.**, Zachariah, O., Odusami, M. O., & Ogunleye, B. (2023). Prediction of stroke disease with demographic and behavioural data using random forest algorithm. *Analytics*, 2(3), 604-617.
7. **Shobayo, O.**, Saatchi, R., & Ramlakhan, S. (2022). Infrared thermal imaging and artificial neural networks to screen for wrist fractures in paediatrics. *Technologies*, 10(6), 119.
8. Ramlakhan, S., Saatchi, R., Sabir, L., Singh, Y., Hughes, R., **Shobayo, O.**, & Ventour, D. (2022). Understanding and interpreting artificial intelligence, machine learning and deep learning in emergency medicine. *Emergency Medicine Journal*, 39(5), 380-385.
9. Ramlakhan, S. L., Saatchi, R., Sabir, L., Ventour, D., **Shobayo, O.**, Hughes, R., & Singh, Y. (2022). Building artificial intelligence and machine learning models: a primer for emergency physicians. *Emergency Medicine Journal*, 39(5), e1-e1.

Conference

1. **Shobayo, O.**, Saatchi, R., Reed, C., & Ramlakhan, S. (2023, January). Correlation of skin temperature with time since injury in paediatric wrist injuries: An infrared thermal image analysis. In *Proceedings of the Annual British Conference on Non-Destructive Testing* (Vol. 2023, No. 1, pp. 1-11). The British Institute of Non-Destructive Testing.

Table of contents

Declaration	i
Abstract	ii
Acknowledgement.....	iii
Dedication	iv
Publications	v
Table of contents	vi
List of Tables.....	x
List of Figures	xi
List of Acronyms and their meaning.....	xiv
List of symbols and their meanings.....	xvii
Chapter 1 – Introduction	1
1.0 Background of the study	1
1.1 Aim and objectives.....	4
1.2 Study’s contributions to the existing knowledge	5
1.3 Outline of the thesis.....	6
1.4 Summary	7
Chapter 2 – Theoretical Background.....	8
2.1 Bone structure	8
2.2 Bone fracture.....	11
2.2.1 Types of fractures.....	11
2.2.2 Bone fracture healing	13
2.2.3 Stages of bone healing.....	15
2.3 Sprain injury.....	18
2.4 Common imaging techniques for fracture diagnosis.....	18
2.4.1 Conventional radiography	18
2.4.2 Computed tomography	19
2.4.3 Magnetic resonance imaging.....	20
2.4.4 Ultrasonography	21
2.4.5 Thermography as a diagnostic tool.	25
2.5 Summary	32
Chapter 3 – Literature Review	33
3.1 Introduction.....	33
3.1.1 Deep neural network overview.....	33
3.1.2 Deep learning in medical imaging, classification and segmentation.....	34
3.1.3 Challenges in utilising deep learning in the medical field.....	34
3.2 Convolutional neural network.....	37
3.2.1 Literature review findings for CNN	39

3.3 Adaptive neuro-fuzzy inference systems	41
3.3.1 Literature review findings for ANFIS	43
3.4 Recurrent neural network	45
3.4.1 Literature review findings for RNN	47
3.5 Autoencoders.....	48
3.5.1 Literature review findings for Autoencoder	51
3.6 Generative adversarial network.....	53
3.6.1 Literature review findings for GAN.....	55
3.7 U-Net.....	56
3.7.1 Literature review findings for the U-Net.....	59
3.8 Transfer learning	60
3.8.1 Literature review findings for transfer learning	62
3.9 Vision transformers	64
3.9.1 Literature review findings for vision transformers.....	65
3.10 Hybrid models.....	67
3.10.1 Convolutional-based hybrid models.....	67
3.10.2 Convolution-transformer based hybrid models	68
3. 11 Discussion	69
3.12 Summary	71
Chapter 4 – Methodology.....	73
4.1 Data collection settings	73
4.2 Study ethics approval	73
4.3 Other processes of ethical considerations	73
4.3.1 Informed consent.....	74
4.3.2 Welfare of participants and adverse event.....	74
4.3.3 Confidentiality of patients.....	74
4.3.4 Data storage and protection.....	74
4.4 Recruitment process	75
4.4.1 Target group	75
4.4.2 Inclusion and exclusion criteria.....	75
4.4.3 Recruitment procedure	75
4.5 Data collection process.....	76
4.5.1 Key considerations in data collection development	76
4.5.2 Data collection protocol	77
4.6 Research equipment, software and image processing	78
4.6.1 Thermal camera and software	78
4.6.2 Analysis software	79
4.6.3 Selecting the ROI and image processing.....	79

4.7 Machine learning and deep learning analysis.....	80
4.7.1 Feature extraction.....	81
4.7.2 Normalisation.....	82
4.7.3 Cross validation.....	82
4.7.4 Image augmentation.....	82
4.8 Summary.....	82
Chapter 5 – Correlation of Skin Temperature with Time Since Injury in Paediatric Wrist Injuries: An Infrared Thermal Image Analysis.....	84
5.1 Introduction.....	84
5.2 Chapter related methods.....	84
5.2.1 Participants’ details.....	85
5.2.2 Image recording.....	85
5.2.3 Region of interest (ROI) segmentation.....	85
5.3 Results.....	86
5.4 Summary.....	90
Chapter 6 – Infrared Thermal Imaging and Artificial Neural Networks to Screen for Wrist Fractures in Paediatrics.....	92
6.1 Introduction.....	92
6.2 Chapter related methodology.....	92
6.2.1 Evaluation statistics.....	92
6.2.2 Selection of region of interest and tracking algorithm.....	94
6.2.3 ROI feature extraction.....	96
6.2.4 Discrimination using multilayer perceptron neural network.....	97
6.3 Result and discussion.....	101
6.3.1 Feature analysis.....	101
6.3.2 Multilayer perception discrimination result for investigation A.....	103
6.3.3 Multilayer perceptron discrimination results for investigation B.....	106
6.4 Summary.....	108
Chapter 7 – Convolutional Neural Network to Classify Infrared Thermal Images of Fractured Wrists in Paediatrics.....	109
7.1 Introduction.....	109
7.2 Chapter related methodology.....	111
7.2.1 Participants’ recruitment and data recording.....	111
7.2.2 Image augmentation.....	112
7.2.3 CNN-based deep learning architecture.....	114
7.2.4 Evaluation metrics.....	117
7.3 Results.....	118
7.3.1 CNN-based deep neural network with augmentation and without dropout.....	119
7.3.2 CNN-based deep neural network without augmentation and with dropout.....	121

7.4 Summary	121
Chapter 8 – Adaptive Neuro-Fuzzy Inference System (ANFIS) Framework for Paediatric Wrist Injury Classification.....	122
8.1 Introduction.....	122
8.1.1 ANFIS Model development for fracture prediction.	123
8.2 Chapter related methodology	123
8.2.1 ANFIS model development.....	123
8.2.2 Cross validation.....	126
8.3 Results and discussion.....	127
8.3.1 Experiment 1: Gaussian membership function based on random normal centre values.	130
8.3.2 Experiment 2: Centre value based on K means clustering for the generalised Membership Function.....	135
8.3.3 Experiment 3: Centre value based on fuzzy c-means clustering for the generalised membership function.....	139
8.4 Summary	144
Chapter 9 – Conclusion and Further work	146
9.1 Conclusion.....	146
9.2 Study’s findings and recommendations	146
9.2.1 Correlation of skin temperature and time since injury	146
9.2.2 Multilayer perceptron based IRT image classification model.....	147
9.2.3 Convolutional neural network based IRT image classification model.....	148
9.2.4 Adaptive Neuro-fuzzy Inference system based IRT image classification model.....	148
9.2.5 Limitations of the study.....	149
9.3 Future work	149
References	151
Appendices	182
i. Patient information for carer	182
ii. Patient information sheet for 5-year-old	187
iii. Patient information sheet for 6 – 12 years old.....	188
iv. Patient information sheet for 13 – 15 years old.....	190
v. Assent form	193
vi. Consent form.....	194
vii. Epigeum ethics training certificate.....	195
viii. Study Questionnaire.....	196
ix. Sample of patients’ anonymised data.....	197
x. Camera details.....	198

List of Tables

Table 3.1. Summary of ANFIS technique for medical image analysis	44
Table 3.2 Recurrent neural network applications.....	48
Table 3.3. Autoencoder based techniques.	50
Table 3.4. GAN based techniques.	54
Table 3.5. U-Net Segmentation techniques.	58
Table 3.6. Summary of transfer learning techniques.....	62
Table 5.1. Mean and standard deviation of time since injury for fracture and sprain.....	88
Table 5.2. Mean and standard deviation of temperature difference representing the ROI (ΔT). 88	
Table 6.1. Average differences between the injured and contralateral uninjured (reference) wrists for the statistical measures, the percentage difference and number of participants differentiated by the statistical measures (fracture (f), sprain (s)).	101
Table 6.2. Investigation A – averaged multilayer perceptron outputs for participants in the test file (averaged over 100 trials). The values have no units.....	104
Table 6.3. Averaged multilayer perceptron outputs (over 100 trials) for participants in the test file, investigation B. The values have no units.....	106
Table 7.1. Layer names and their learnable parameters.	114
Table 7.2. Hyperparameters for CNN based deep neural network.....	115
Table 7.3. Number of images used for CNN training and validation.....	119
Table 7.4. Evaluation metrics for CNN-based deep neural network with augmentation and without dropout.....	119
Table 7.5. Evaluation metrics for CNN-based deep neural network without augmentation and without dropout.....	121
Table 8.1. Summary of performance for training and validation targets with epochs(t) for experiment 1	130
Table 8.2. Summary of performance for training and validation targets with epochs(t) for Experiment 2.....	135
Table 8.3. Summary of performance for training and validation targets with epochs(t) for Experiment 3.....	139
Table 8.4. Comparison of evaluation metrics.....	143

List of Figures

Figure 2.1. A schematic diagram of the compact and cancellous bone. A represents the longitudinal section of the bone, (B) depicts a magnified view of the compact bone and (C) shows a section of the flat bone.....	9
Figure 2.2. Osteoblasts and Osteocytes formation	11
Figure 2.3. Showing different types of fracture. (A) represents Greenstick fracture (B) represents oblique fracture (c) represents comminuted fracture (D) represents transverse fracture (E) represents compound or open fracture and (F) represents compression fracture, with the arrow showing the direction of the cancellous bone collapsing on itself	12
Figure 2.4. Dorsal to volar (A – F) CT scan of a fractured wrist area.....	20
Figure 2.5. Showing US (A) and X-ray (B) images of a patient with wrist fracture. The fracture site is denoted by the red arrow in the US image.	22
Figure 2.6. Electromagnetic Spectrum showing the IR band.	28
Figure 2.7. A. IR thermography helping in the diagnosis of pregnancy in a rhinoceros. B. IR thermographic image showing inflamed hindleg patellar damage in a horse. C. IR thermal image helping to diagnose fracture in the right tibia of a toddler. D. IR thermal image helping to detect wrist fracture in children.....	29
Figure 2.8. Active thermography set up.	30
Figure 2.9. Passive Thermography.	31
Figure 3.1. A summary of deep learning usage with medical images.....	34
Figure 3.2. An example of CNN used for the analysis of infrared thermal images	37
Figure 3.3. ANFIS Basic architecture	42
Figure 3.4. Structure of the RNN.	46
Figure 3.5. Structure of the GAN network	53
Figure 3.6. The architecture of U-net.	57
Figure 3.7. Structure of a transfer learning model – AlexNet.	61
Figure 3.8. Vision Transformer architecture.	64
Figure 4.1. Depicting the successful tracking of the ROI	77
Figure 4.2. A schematic representation of the initial data collection protocol.	78
Figure 4.3. Showing the screenshot of the GUI of the FLIR Research Max 4 software that controls the FLIR T630sc camera	79
Figure 4.4. Showing the diagrammatic representation of the ratios used in determining the ROIs for the wrist and arm regions.	80
Figure 5.1. Histograms of ΔT for (a) fracture, (b) sprain.	87

Figure 5. 2. Histograms of time since injury for (a) fracture, (b) sprain.	87
Figure 5. 3. Plot of ΔT against TSI for fracture.	88
Figure 5. 4. First phase of plot in Figure (5.3).	89
Figure 5. 5. The second phase of plot in Figure (5.3).	89
Figure 5. 6. Plot of ΔT against TSI for sprain	90
Figure 6.1. Statistical measures used to analyse effectiveness of MLP and IRTI to differentiate wrist fracture and wrist sprain	93
Figure 6.2. IR thermal images indicating the region of interest as a dotted line: (a) fractured left wrist; (b) uninjured right wrist.	94
Figure 6.3. IR thermal images indicating the region of interest as a dotted line: (a) sprained left wrist; (b) uninjured right wrist.	95
Figure 6.4. A typical averaged region of interest	95
Figure 6.5. The region of interest of figure converted to cells of 10×10 pixels, and each cell represented by its mean temperature value.	96
Figure 6.6. The multilayer perception used to differentiate wrist fracture and wrist sprain ...	98
Figure 6.7. Boxplots of average differences between the injured and uninjured contralateral wrists for the statistical measures. Note: the label for vertical axis is not shown, as the variables do not have the same unit. The unit for minimum, maximum, mean, standard deviation (std) and interquartile range (IQR) is °C. Skewness and kurtosis have no units.	102
Figure 6.8. The receiver operating characteristic curve (ROC) for investigation A	105
Figure 6.9. Plot of averaged multilayer perceptron outputs (over 100 trials) for participants in the test file in investigation A. The stars and circles represent participants with fracture and sprain, respectively. The dashed line is the threshold.	106
Figure 6.10. The receiver operating characteristic curve (ROC) for investigation B.	107
Figure 6.11. Plot of averaged multilayer perceptron outputs for participants in the test file for investigation B. The stars and circles represent participants with fracture and sprain, respectively. The dashed line is the threshold	108
Figure 7.1. Operations to identify wrist fracture	111
Figure 7.2. Infrared image of a participants' hands: (a) fracture of left wrist; (b) sprained left wrist. The region of interest is shown by the blue dotted line.	112
Figure 7.3. Magnitude frequency spectra of a typical region of interest for (a) fractured wrist ROI and (b) sprained wrist ROI.	113
Figure 7.4 CNN-based deep neural network used in the study.	115

Figure 7.5. The CNN model training with validation data. The blue graph represents the model's training, and the black graph represents the validation using the validation dataset with validation carried out after 250 epochs.....	117
Figure 7.6. The CNN model loss with validation data. The red graph represents the model training loss (or error), and the black graph represents the validation error using the validation dataset with validation carried out after 250 epochs.	118
Figure 7.7. Confusion matrix for CNN-based deep neural network for wrist fracture versus sprain. The squares contain the number of images for validation dataset, total: 3588 images.	120
Figure 7.8. Receiver operating characteristic curve for CNN-based deep neural network.	120
Figure 8.1. 3-input 2-output ANFIS architecture of the fracture classification system.	125
Figure 8.2. Training and Validation MSE for Experiment 1	131
Figure 8.3. Confusion Matrix of Experiment 1	132
Figure 8.4. AUCROC of Experiment 1	133
Figure 8.5. Contour Plot of Experiment 1	134
Figure 8.6. Training and Validation MSE for Experiment 2.....	136
Figure 8.7. Confusion matrix of Experiment 2.....	137
Figure 8.8. AUCROC for Experiment 2.....	137
Figure 8.9. Contour Plot of Experiment 2	138
Figure 8.10. Training and Validation MSE for Experiment 3.....	140
Figure 8.11 Confusion Matrix for Experiment 3.....	141
Figure 8.12. AUCROC for Experiment 3.....	141
Figure 8.13. Contour plot for Experiment 3	142
Figure 8.14. Evaluation metrics across the experiments.	143
Figure 8.15 Accuracy comparison with other models.....	144

List of Acronyms and their meaning

3D – 3-Dimensional
A&E – Accident & Emergency
AD - Alzheimer's Disease
AI – Artificial Intelligence
ANFIS – Adaptive Neuro-Fuzzy Inference System
ANN – Artificial Neural Network
AO – Arbeitsgemeinschaft für Osteosynthesefragen
AP – Anteroposterior
AUC – Area Under Curve
AUC – Area Under Curve
BCE – Binary cross Entropy
BGRU – Bi-directional Gated Recurrent Unit
BMD – Bone Mineral Density
BraTs – Brain Tumour Segmentation
BSbFWT – Bayes Shrinkage Fused Wavelet Transform
CAD – Computer Aided Diagnostics
CGAN – Cycle GAN
CLA – Cervical lymphadenopathy
CMA – Convolutional Mesh Autoencoder
CNN – Convolutional Neural Network
CPU – Central Processing Unit
CT – Computed Tomography
DCGAN – Deep Convoluted
DCNN – Deep CNN
DCNN – Deep Convolutional Neural Network
DeiT – Data Efficient Image Transformer
DFT – Discrete Fourier Transform
DL – Deep Learning
DNA – Deoxy Ribonucleic Acid
DNN – Deep neural Network
DRR – Digitally reconstructed Radiographs
DXA – Dual energy X-ray Absorptiometry
EEG – Electroencephalogram

ESE – Entrant Skin Exposure
FC – Fully Connected
FCM – Fuzzy C-Means
FFT – Fast Fourier Transform
FID – Fréchet Inception Distance
FLIR – Forward Looking Infrared
FLoP – Floating Point of Operation
FN – False Negative
FNR – False Negative Rate
FP – False Positive
FPR – False Positive Rate
GAN – Generative Adversarial Network
GPU – Graphic Processing Unit
Grad-CAM – Gradient-weighted Class Activation Mapping
GRU – Gated Recurrent Unit
GUI – Graphical User Interface
HBA – Honey Badger Algorithm
HPC – High Performance Computing
hrMRI – high-resolution Magnetic Resonance Imaging
HSSA – Hybrid Salp Swam Algorithm
IQR – Interquartile range
IR – Infra-Red
IRTI – Infra-red Thermal Images
KHO – Krill Herd Optimisation
LSTM – Long Short-Term Memory Network
MAE – Mean Absolute Error
MATLAB – Matrix Laboratory
mCT – micro-Computed Tomography
MF – Membership Function
MF – Membership Function
MFLA – Multi-level Feature Assembly
ML – Machine Learning
MLFA – Multi-level Feature Assembly
MLP – Multi-Layer Perceptron

MRI – Magnetic Resonance Imaging
MSE – Mean Squared Error
NHS – National Health Service
NoP – Number of Parameters
NPV – Negative Predictive Value
OI – Osteogenesis Imperfecta
OTA – Orthopedic Trauma Association
PD – Parkinson’s Disease
PNSR – Peak Signal to Noise Ratio
PPV – Positive Predictive Value
PSO – Particle Swarm Optimisation
PSO – Particle Swarm optimisation
PVT – Pyramid Vision Transformer
QCT – Quantitative Computed Tomography
R-CNN – Recurrent CNN
rEBUS – Radial Endobronchial Ultrasound
REID – Risk Exposure Induced Deaths
ReLU – Rectified Linear Unit
ResNet – Residual Neural Network
RF – Radio Frequency
RMSE – Root Mean Square Error
RNN – Recurrent Neural Network
ROC – Receiver Operating Characteristics
ROC – Receiver Operating Characteristics.
ROI – Region of interest
SAE – Stacked Autoencoder
SC – Syndromic Craniosynostosis
SD – Standard deviation
SGD – Stochastic Gradient Descent
SPSS – Statistical Package for Social Science
SRNet - Stack Multi-Connection Simple Reducing Net
SSIM – Structural Similarity Index Metric
SVM – Support Vector Machine
Swin – Shifted Window Transformer

TL – Transfer Learning
TN – True Negative
TNT – Transformer in Transformer
TP – True Positive
TSI – Time since Injury
UCI – University of California Irvine
US – Ultrasound
USB – Universal Serial Bus
VGG – Visual Geometric Group
ViT – Vision Transformer
WGAN – Wasserstein Generative Adversarial Network
YOLO – You Only Look Once

List of symbols and their meanings

ΔT – Change in Temperature
 $^{\circ}\text{C}$ – Degree Celsius
MHz – Mega Hertz
 μm – Micrometre
 $^{\circ}\text{K}$ – Degree Kelvin
 ε – Emissivity
% – Percentage
 \leq – Less than or equal to
fps – Frames per seconds
nm – Nanometre
mm – Millimetre

Chapter 1 – Introduction

In this chapter the background and purpose of this PhD study, its aim and objectives, its study's contributions and a brief overview of the following chapters are provided.

1.0 Background of the study

The term bone fracture is related to any medical condition that deals with discontinuity of a bone partially or a complete breakage. A bone fracture could be caused by a medical condition (e.g. osteogenesis imperfecta) or an accident such as a fall. In extreme cases, fracture may result in the bone breaking into multiple pieces. When the bone is subjected to continuous stress and constant compression, the bone tissue's ability to withstand stress diminishes over time. When the bone tissue has reached its structural limits, it begins to fracture (Oryan, et al., 2013). A sprain on the other hand relates to damaged ligaments due to an injury. Most damaged ligaments are caused by injuries that forces the joint beyond its range of motion. Sprain occurs mostly where two or more bones are connected, with majority of sprain injuries occurring in the ankle, wrist and knee. Sprains are typically less severe as compared to fractures with symptoms including pain, increased local temperature and swelling.

Bone fracture is a common cause of injury, and it has accounted for high medical costs reaching billions of pounds worldwide. A large proportion of this cost is accrued for treating the injury and other costs can be attributed to loss of productivity in the workplace (Ghiasi et al., 2017). Further costs come as result of x-ray radiograph taken from suspected fracture leading to confirmation of the injury type, i.e., fracture or sprain. There are also the radiology staff cost and related equipment cost and radiologist's time. (Yousefifard et al., 2016) reported that the main modalities for fracture diagnosis are ultrasonography and radiography, with either of them suitable for detecting fractures in different parts of the body, with X-ray radiography being the gold standard. These modalities have their fallbacks as ultrasound images for instance are quite noisy (speckle noise) making visualisation of some fracture types difficult while radiography (X-ray) has proven to be a cause of DNA damaging carcinogen. Although X-ray images are still the gold standard for the detection of bone fractures, they have shown to have multiple rates of abnormalities especially for the diagnosis of long bones. This has led to X-ray producing high miss rate (Chai, 2011). In toddler's fracture, i.e. fracture of tibia in children aged below 3 years, X-ray scan may not detect the fracture on the first scan. Therefore, the child's leg is put in plaster,

and a second X-ray scan is taken around 10 days later and then if there is a fracture it could be visible on the X-ray. This creates stress for the child and the need for repeat X-ray examination.

Children are highly prone to bone fractures due to injuries sustained at playgrounds and they account for 25% of injuries and accidents encountered at hospital's A&E department and the most common type of fracture reported is fracture of the distal radius. The degree of fracture, determined through a hospital triage will determine if the child will require an X-ray scan. The main reason for the extremes is (Seens et al., 2021):

- i. Pathological bone conditions.
- ii. Accidental trauma.
- iii. Non-accidental trauma.

A pathological bone fracture can be described as a condition when the bone tissue has undergone remodelling by an internal process. In children, this can be caused by several varied conditions from tumours to infections and neuromuscular diseases. Canavese et al. (2016) grouped some of the pathological undertone to bone fracture as focal, diffuse and neuromuscular. Some of the conditions mentioned under focal are congenital pseudarthrosis of tibia, Osteomyelitis, and Unicameral bone cyst. Conditions categorised as diffuse are Osteogenesis imperfecta, Enchondromatosis and Osteopetrosis. Accidental trauma causes of bone fracture extremities as mentioned by (Seens et al., 2021) include accidents caused by motor vehicles, falls (from a height or while running) during fights and recreational activities like riding a bike. Non-accidental trauma fractures refer to bone injuries that occur as a result of intentional harm rather than accidental causes, most commonly associated with child abuse or inflicted injury. These fractures are not due to typical activities such as falls or sports but arise from deliberate force applied to the body. Some examples of non-accidental trauma include fractures caused by physical abuse, such as hitting or shaking a child, rib fractures resulting from forceful squeezing, long bone fractures (e.g., femur or humerus) in infants who are not yet walking, and skull fractures caused by intentional impact. The prevalence of the injury means that children are posed to a high-risk exposure to the current means of diagnosis which could be X-ray.

X-ray generated from a radioactive material occurs naturally and can also be technically produced. They have found usage in different medical procedures such as diagnostic radiology, radiographic procedures for therapy and nuclear medical diagnostics. The levels of dose (exposure) to the ensuing radiation vary with different medical interventions. The general dose used in the medical field is typically at 1.7 mSv but is significant when compared to the dose of

4.0 mSv which the safe limit for any human exposure to radiation. As with the many medical use of radiology for radio therapy also come with the inherent dangers of radiation. Research has shown that children are more at risk of developing carcinogenic malformations when exposed to X-rays than adults (Reed et al., 2020). X-rays have been linked to molecular and atomic damages to biological tissues especially in children (Buchberger et al., 2022). For example, epidemiologic studies were carried out by (Preston et al., 2007) to investigate the incidence of neoplasms and haemoblastoma, which are different forms of cancers among the survivors of the Hiroshima bombings in Japan, linking the occurrence of the disease to the distance from the bombing epicentre with people closest to the epicentre having a higher dose of the radiation. REID values which stand for Risk Exposure induced Cancer deaths has also been used as a yard stick to measure the dosage of radiation to the cause of cancer (Nahangi & Chaparian, 2015). With Entrant Skin Exposure (ESE) values of children that have undergone pelvis and chest X-ray examinations, female children had the highest REID values (likelihood to develop cancer related diseases). In children, the required dose of radiation depends on age, gender, and how much a dose a particular organ has been exposed to (Shatsky, 2017). Females are at a higher risk to stochastic health effects when compared males and children between the ages of 10-14 (Shatsky, 2017) (Nahangi & Chaparian, 2015). Neonatal, i.e., unborn babies have been found to have no risk of exposure (Shu et al., 1994).

Therefore, there is a need to provide a screening tool that reduces the number of unnecessary X-ray scans taken to diagnose bone fractures. Currently, the use of Infra-red (IR) thermal imaging has been growing in the field of medical diagnosis. A modern IR thermographic camera very accurately produces images by measuring the temperature produced by the infrared radiation emanating from surfaces including the human skin as a temperature map. Another useful feature of the technology is its noncontact procedure, allowing the measurement of objects that can be considered inaccessible and dangerous (Owen et al., 2017) (Kastberger & Stachl, 2003). According to (Owen et al., 2017), IR thermometry has seen an increased use in treating sport injuries, diagnosing organs to get an insight into the severity of some diseases like cancer and arthritis. Other areas where thermographic images have been used in medical diagnosis especially for bone fractures are diagnosing of acute limp injury in children (Owen et al., 2017), paediatric wrist fractures (Reed et al., 2020), evaluation of the effect of vertebrae fracture in children with osteogenesis imperfecta (Fane De Salis et al., 2018). From the thermal images produced, a statistical analysis showed significant temperature difference in the fracture site, considered as region of Interest (ROI), due to an increase in the flow of blood towards the area. (Reed et al.,

2020) were able to differentiate temperature changes between fractures and sprain, i.e., non-fractures. They showed that fractures caused about 1.519% increase in the mean skin surface temperature while sprains caused about 0.971% temperature increase when compared to the temperature an uninjured wrist acting as reference.

This study aimed to develop artificial intelligence-based approaches, including deep learning, to automate the differentiation of wrist fracture and sprain in children using infrared thermal images and thus develop as a new tool for fracture screening. In an earlier study (Reed et al., 2020), 105 children with wrist injury were both x-rayed and thermal imaged. Roughly 50% of the cases had wrist fracture. This database was used to carry out the study. Therefore, there was a suitable database for the purpose of this study (required ethics granted). The study thoroughly explored wrist injury as it is a more common type of fracture in children. The manner the time since the injury affects the temperature of the injured wrist region in fracture and sprain was also studied.

1.1 Aim and objectives

The aim of this study was to develop and evaluate artificial intelligence-based methods to process and interpret paediatric infrared thermal (IRT) images to assist with wrist fracture screening.

The objectives were:

- i. Devise suitable IRT imaging preprocessing and feature extraction to facilitate differentiation of wrist fracture from sprain.
- ii. Investigate the correlation of skin temperature with time since injury within the fracture and sprain data of the subjects.
- iii. Investigate the efficacy of a shallow neural networks with manual feature extraction in the classification of IRT images of wrist fractures and sprain.
- iv. Investigate the complexities of deep neural networks with convolutional filters for feature extraction in improving the classification metrics of shallow networks like the MLP in objective (iii).
- v. Evaluate the efficiency of using a lightweight, fuzzy rules based explainable neural network algorithm such as ANFIS to classify the thermal images of the fracture and sprain.

The main research challenge has been to identify the most effective techniques for the classification of infrared thermal images of the fractured and sprained wrists. Also, developing a deep learning model with better evaluation metrics.

1.2 Study's contributions to the existing knowledge

The contributions of this study aligned with the objective in section 1.1 is listed below:

- i. A quantification of the relationship between skin temperature and time since injury (TSI) in paediatric wrist injuries using infrared thermal imaging was undertaken. In the 40 patients (19 fractures and 21 sprains), there was evidence that fractures exhibited a biphasic temperature profile: an initial 3.3-hour phase with no significant correlation between the TSI and skin temperature followed by a moderate positive correlation between TSI and skin temperature with ($r = 0.680$) as inflammation progresses. Sprains however showed no significant correlation between TSI and skin temperature, and fractures produced a more consistent temperature rise. These findings offer novel insight into the temporal evolution of inflammatory responses and suggest infrared thermography could serve as a non-invasive tool for monitoring fracture healing and differentiating fractures from soft-tissue injuries in clinical practice and research contexts.
- ii. The integration of infrared thermal (IRT) imaging and multilayer perceptron (MLP) artificial neural network (ANN) for paediatric wrist-fracture screening. A grid-based region of interest (ROI) feature extraction method that divides IRT images into 10×10 -pixel cells was devised. Then 50 highest-temperature regions were selected, and their statistical descriptors (e.g., kurtosis, standard deviation, and interquartile range) were computed. These were processed by the MLP thereby capturing spatial temperature heterogeneity at injury sites. Template-matching tracking realigns ROIs across sequences, and image averaging reduced thermal noise, enhancing measurement reliability. To address limited sample sizes (19 fractures, 21 sprains), two investigations employing repeated random train-test splits over 100 trials reduced the selection bias and yielded sensitivity (84.2%) and specificity (71.4%). This proof-of-concept highlighted IRT imaging – MLP's potential to reduce unnecessary radiography and streamline emergency triage. The methodological contributions such as innovative ROI segmentation, feature selection, and small-sample validation strategies helped lay the groundwork for follow-on research to refine classifier architectures, expand to other fracture types and adult populations; integrate real-time IRT imaging screening in clinical workflows.
- iii. A bespoke convolutional neural network (CNN) was designed to classify paediatric wrist fractures from infrared thermal images, advancing beyond prior manual-feature methods by leveraging deep learning for automatic feature extraction. Key aspects include inclusion of fast Fourier transform (FFT) of injury-site ROIs to suitably represent the

image features, followed by image augmentations, which constituted random rotations ($\pm 90^\circ$), translations (± 3 pixels), and shear (± 2) in order to enrich the training set and mitigate overfitting in a small sample of 11,960 images from 40 participants. Two model variants were critically compared: one employing augmentation without dropout, achieving 76% accuracy, 88% sensitivity, 68% specificity, and an AUC of 0.82; and another using a 0.2 dropout layer without augmentation, which underperformed (58% accuracy, 50% sensitivity). These findings demonstrated that tailored augmentation strategies can substantially improve CNN generalization on limited medical-imaging datasets. Although dropout can be very effective in helping CNN models to reduce overfitting, where datasets is depleted, this might not be the case.

- iv. A tailored Adaptive Neuro-Fuzzy Inference System (ANFIS) for classifying paediatric wrist fractures from infrared thermal-image descriptors was devised. It merged the interpretability of fuzzy “IF-THEN” rules with artificial neural-network learning to yield both transparency and nonlinear modelling power. By selecting only three statistical thermal features which are standard deviation, interquartile range, and kurtosis from the IRT image features, this work demonstrated that a compact input set can achieve reliable discrimination, streamlining model complexity without sacrificing accuracy. A rigorous five-fold cross-validation and grid search over Gaussian-membership-function widths and fuzzy-rule counts ensured an optimal generalisation with early-stopping safeguarding against overfitting. Three premise-parameter initialisation strategies which included random Gaussian seeds, K-means clustering, and fuzzy C-means clustering, revealed that K-means based seeding accelerates convergence (48 s versus 194 s for random), stabilised validation error and attains balanced sensitivity (67%) and specificity (80%) with an area under receiver operating curve of 0.87, outperforming both MLP and CNN alternatives. These findings highlighted the pivotal role of data-aware cluster placement in neuro-fuzzy model design.

1.3 Outline of the thesis

The outline of the thesis is presented as below:

Chapter 2: Theoretical background. In this chapter, studies related to the diagnosis of wrist fractures using several methods is presented.

Chapter 3: Review of Artificial neural network techniques for medical image analysis. In this chapter, the review of several artificial neural network techniques to classify medical images was presented.

Chapter 4: Methodology. This chapter presents the research problem, type of data, the data collection method, the subjects' details, and ethical considerations.

Chapter 5 Investigation of skin temperature with time since injury in paediatric wrist injuries. In this chapter, the correlation of skin temperature at the site of the wrist fracture or sprain with time since the occurrence of injury is investigated.

Chapter 6: Investigation of shallow neural network with manual feature extraction. The suitability of shallow neural network such as MLP with manual feature extraction to distinguish between IRT images of wrist fracture and sprain is investigated.

Chapter 7: Investigate the complexities of deep neural networks with automatic in improving the classification metrics. Deep learning algorithm such as CNN with automatic feature extraction to improve classification metrics of the IRT images of wrist was investigated.

Chapter 8: Investigating the use of Adaptive Neuro-Fuzzy Inference System (ANFIS) Framework for Paediatric Wrist Injury Classification. Fuzzy based neural network architecture was investigated to classify the IRT images of the subject to explore the lightweight and explainability that it provides.

Chapter 9: Conclusion and further work. The study is summarised, and further research work is presented.

1.4 Summary

In this chapter, the concept of bone injuries which include fracture and sprain in paediatrics were introduced. X-ray has been the gold standard technique for diagnosing bone injuries including wrist injuries but has some issues ranging from the taking of unnecessary x-ray radiographs for sprain patients, leading to unnecessary cost incurred by the NHS. There is also the inherent negative aspect of exposure to X-ray radiation. IRT imaging has been emerging as an alternative tool in medical diagnosis. It records skin temperature by measuring the amount of IR emissions. The statistically significant difference in temperature between wrist fracture and sprain injuries can be leveraged by IRT imaging to train artificial intelligence models to be able to distinguish between fracture and sprain, hereby reducing the amount of X-ray needed, which will in turn save costs and limit overexposure to radiation.

Chapter 2 – Theoretical Background

In this chapter, the related theories surrounding bone fractures and the healing process were presented. Several imaging techniques to visualise the type of bone fracture that occurred were also discussed. Finally, a look into Infrared thermal imaging (IRT) techniques as they grow in relevance as a diagnostic tool for disease detection especially fracture and how artificial intelligence became part of the diagnostic process.

2.1 Bone structure

The bone is a very important part of the human anatomy, and it is made up of a bony skeleton with the main function of providing strong support and optimal mechanical structure for ease of movement. It also provides support for soft tissues and muscles of the body (Ghiasi et al., 2017). The structural property is made up of water (20%) and dry components of organic and inorganic substances (comprising 30%-35% and 65%-75% respectively) such as collagen, hydroxyapatite, proteoglycans (small amounts) and non-collagenous proteins. The rigidity of the bones and its compression is made possible by the inorganic compounds, while the tensile strength of the bone is made possible by the organic components. The quantities at which this is present in the human body is dependent of the following demographics such as race, sex, age, medical history of the bone and its anatomical position. The skeletal system provides a shield for the marrow and shield most of the organs in the abdomen/thoracic area of the body. Its calcium compounds consumed from our food is get deposited on the bone and also forms a major part of the process of manufacturing of blood cells, a process known scientifically as haematopoiesis (Yousefifard et al., 2016).

Figure 2.1 (Wallig et al., 2017) describes the overview of the bone section. Bone tissues can be classified into two types namely: Cancellous bone (trabecular) and Cortical bone (compact). This classification is made with reference to the porosity of the bones whereby the trabecular has porosity between 50-95 percent, while the cortical bone has a porosity of about 5-10 percent.

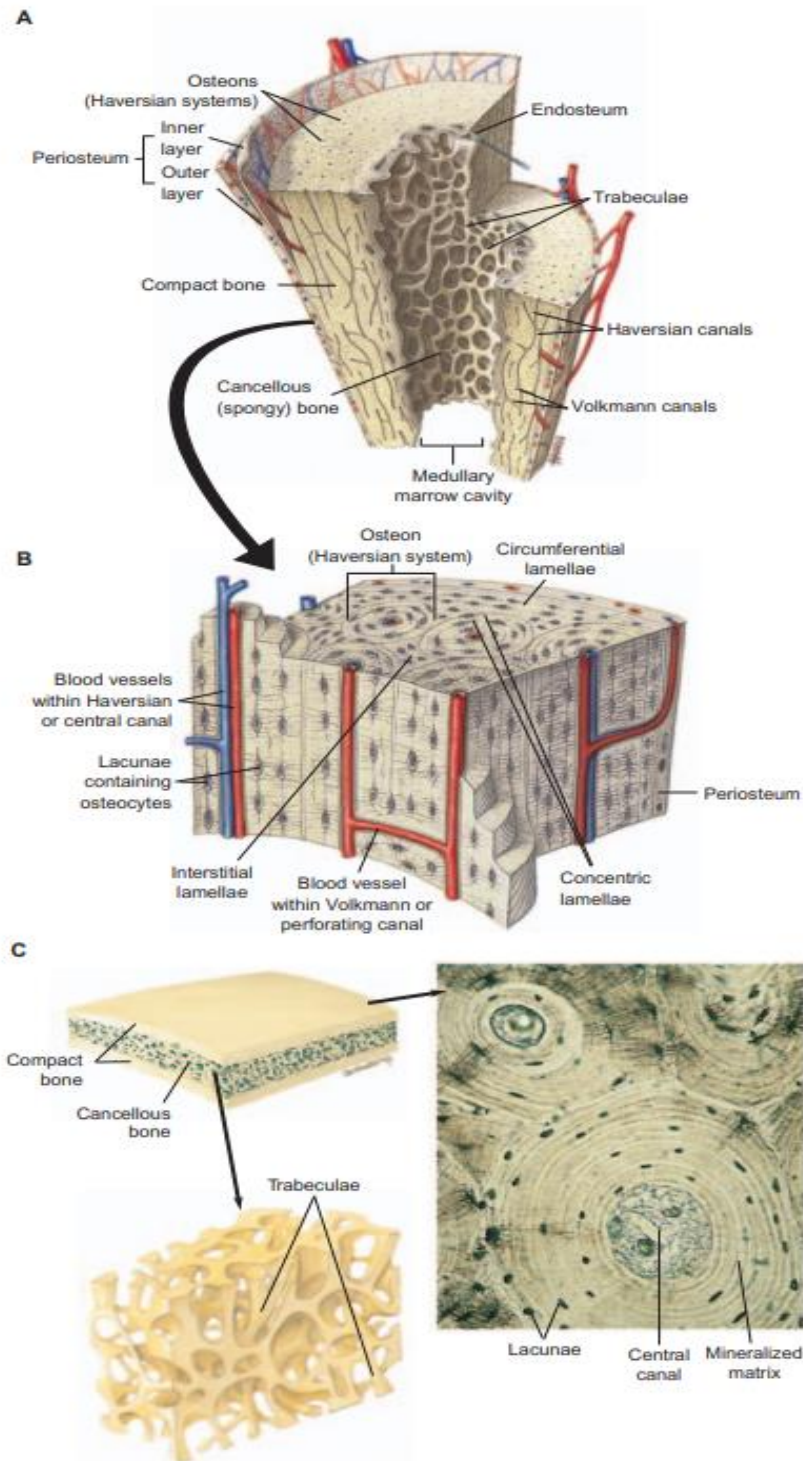


Figure 2. 1. A schematic diagram of the compact and cancellous bone. A represents the longitudinal section of the bone, (B) depicts a magnified view of the compact bone and (C) shows a section of the flat bone (Wallig et al., 2017).

Examples of bones that possess the characteristics of the trabecular are the flat bones, cuboidal bones, and the distal parts of long bones. The pores of the bone tissues are connected to each other and contains the substance known as marrow. The contents of the marrow are significant

to the support mechanism provided by the bones as this deals with regenerative ability and the health of the bone. The nerves, blood vessels and different types of cells, all found inside the marrow are responsible for the production of basic blood cells. The bones can be arranged in different forms of plates and struts, combining to form what is known as the bone matrix. The matrix is about 200µm thick. The arrangement can also be varied (Yousefifard et al., 2016).

Vascular porosity is the most prominent of the cortical bone type with its largest diameter spanning 50µm. They are formed by the alignment of the Haversian canals and Volkmann's canal. They are other porosities with very small diameter spanning a length of 10 nm. Cavities and the gaps found in hydroxyapatite forms the major type of this bone tissues (Ghiasi et al., 2017).

Bones possess the ability to grow, repair itself when fractured in a process known as fracture healing, alter their shape due to external modelling or remodelling and can renew themselves internally (internal remodelling). The process that causes changes in bone structure are catalyst by different sorts of physiological patterns, mechanical stress and strain and hormones. These growth curves occur with different forms of trigger. During childhood, the process of growth and modelling of bones occurs. Throughout life, the process of internal remodelling occurs. This forms the key part of the internal structure of the bone, subsequently determining the mechanical strength of the bone and its ability to naturally repair little damage to bones such a sprain (Chai et al. (2011).

The internal surface of the bone arrangement or matrix provides the avenue for remodelling of bones. This can be seen in the cancellous bone (trabecular surfaces) and the Haversian system provides that of the cortical bone. Alteration to bone structure i.e., modelling or remodelling can only be made possible on these surfaces (Chai, H. Y. et al., 2011).

At the cellular level, the bone can be said to be comprised of osteoblasts, osteoclasts, osteocytes and osteogenic precursor cells (Chai, H. Y. et al., 2011). Osteoblast produces bone and are differentiated from the mesenchymal cells. They are produced by the stromal tissue of the bone marrow, and they mainly contain collagen. Osteocytes on the other hand can be classified as a fully matured osteoblast that remained in the lacunae. They are most often referred to as mesenchymal osteoblast or surface osteoblast by scientist. Osteoclasts on the other hand are bone cells that remove bones. To do this, acids and enzymes are produced to demineralise bones. These

cells are also produced from the bone marrow (Chai, H. Y. et al., 2011). The formation of the osteoblasts and osteocytes is shown in Figure 2 (Wallig et al., 2017).

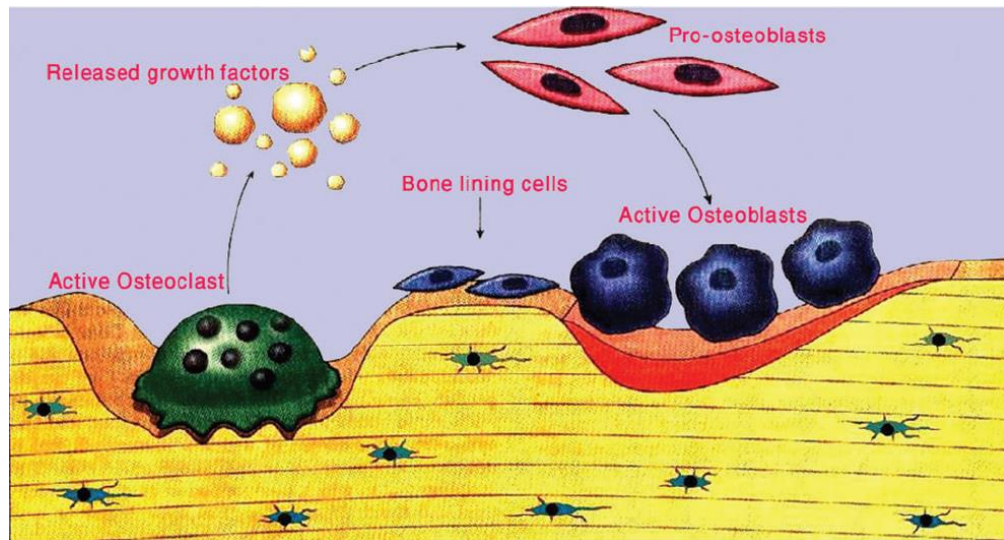


Figure 2. 2. Osteoblasts and Osteocytes formation (Wallig et al., 2017)

2.2 Bone fracture

The term bone fracture is related to any medical condition that deals with discontinuity of a bone either partially or a complete breakage. In other extreme cases, fracture may be as a result of bones breaking into several pieces. When the bone is subjected to continuous stress and constant compression, the bone tissue's ability to withstand strength decreases over time. When the bone tissue has reached its limits, then the bone begins to crack or break, leading to fracture (Chai, H. Y. et al., 2011). Other causes can be related to mild trauma due to patients suffering certain medical conditions that causes bones to be weak. Examples of such conditions are bone cancer, osteogenesis imperfecta, osteoporosis and osteopenia.

2.2.1 Types of fractures

To provide the appropriate treatment plan for fractured bones, especially in cases when there is a significant breakage, it is necessary to streamline the kind of fracture that has occurred. Classification of bone fracture this way, makes life easier for the orthopaedic surgeon. Fractures in bones can be classified based on different characteristics used in describing the fracture. These include causal factors, location of injury, degree of severity of the fracture and morphology (Abakka et al., 2023). Fractures based on causal factors can be presented as oblique (which is a result of bending force, causing the fracture line to be oblique), spiral (the causal indicator is a torsional force, causing the fracture line to run in several planes), transverse (fracture line in

orthogonal to the long axis of bone and is caused by bending forces), and comminuted i.e., caused by severe force impacting directly on the bone making it shatter into different fragments (Abakka et al., 2023).

Other common presentations include crush fracture or compression (occurs when there is compressional force on a spongy bone), gunshot fracture (this occurs when there is an incomplete bone breakage: bending of the inner cortex and breakage of the outer cortex) and avulsion fracture (presented as the detachment of a bone piece from the main bone). In classifications based on the aetiology (causes based on history, attribution or reason), 3 types of fracture can occur. The first is traumatic (due to unexpected excessive force), fatigue (caused by repetitive stress) and pathological (bone weakens due to diseases such as tumours and disuse of bone) (Oryan, Ahmad et al., 2015).

Based on nature of the injury, fractures may also be considered as closed (simple fracture) and open (compound fracture). The open fracture can be further classified into internally and externally open fracture. In either of these cases, there is a direct link between the fractured site and the external environment, which posed infectious risk to the fractured site thereby requiring special attention with the management of the wound on the skin and when surgical procedures are considered. Massive open fractures are considered complicated to handle as it takes time to heal due to double healing process which includes the bone and the skin leading to massive delay in the healing process (Oryan, Ahmad et al., 2015). A cross section of the different types of fracture discussed above is shown in Figure 2.3 (Oryan, Ahmad et al., 2015) .

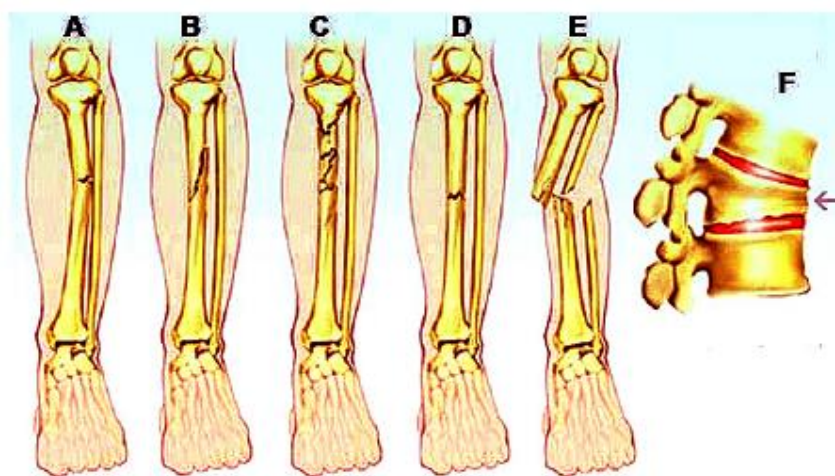


Figure 2. 3. Showing different types of fracture. (A) represents Greenstick fracture (B) represents oblique fracture (c) represents comminuted fracture (D) represents transverse

fracture (E) represents compound or open fracture and (F) represents compression fracture, with the arrow showing the direction of the cancellous bone collapsing on itself (Oryan, Ahmad et al., 2015).

2.2.2 Bone fracture healing

In the development of an effective treatment approach using various biomaterials and therapeutic agents, it is essential to understand the complex processes involved in bone fracture healing. The primary objective of fracture healing is to regenerate mineralised tissue at the fracture site, restoring the bone's mechanical strength and structural integrity to ensure normal functionality. Bone regeneration plays a crucial role in re-establishing the original morphology and structure of the tissue (Marsell & Einhorn, 2011). The process of fracture healing varies based on the type of fracture and the method used for fracture fixation. Additionally, factors such as the choice of biomaterials or grafts for reconstruction and the specific therapeutic agents used can significantly influence the healing pattern and overall response (Tsiridis et al., 2007). However, the precise effects of these treatments are still not quantifiable. From a histological perspective, and depending on the method of fracture fixation, bone healing is classified into two primary categories: primary (direct) fracture healing and secondary (indirect) fracture healing.

These classifications help in understanding the biological mechanisms that guide bone repair and in designing more effective treatment strategies (Marsell & Einhorn, 2011) (Tsiridis et al., 2007).

2.2.2.1 Primary or direct fracture healing

Fixation methods that apply compression and stability at the fracture site enable direct bone healing, also known as primary healing or union. This type of healing occurs when the bone is surgically realigned through open reduction and secured using rigid internal fixation, preventing the formation of gaps between bone fragments. Primary healing can happen in two ways: contact healing and gap healing (Oryan, Ahmad et al., 2015).

In contact healing, the bone ends are closely aligned, and the strain between fragments is minimal. This allows direct remodelling of the cortical bone, a dense outer layer of bone that provides strength. The process involves cutting cones, which are specialised structures formed near the fracture site. These cones contain osteoclasts—cells that break down bone to create longitudinal cavities. These cavities are then filled with new bone by osteoblasts, the cells responsible for bone formation. On the other hand, gap healing occurs when there is a small gap

between the bone fragments. Unlike contact healing, bony union does not occur simultaneously with Haversian remodelling, a process where the bone is reorganised into its normal structure (Marsell & Einhorn, 2011).

However, both processes lead to the formation of Haversian canals, which allow blood vessels to penetrate the bone, facilitating remodelling into strong lamellar bone without forming a callus, the temporary tissue that forms during indirect healing (Marsell & Einhorn, 2011) (Oryan, Ahmad et al., 2015).

2.2.2.2 Secondary or indirect fracture healing

The most common method used in fracture healing is the secondary or indirect healing, which involves intramembranous ossification and endochondral ossification, both leading to the formation of a callus (a temporary repair tissue) (Dimitriou et al., 2005). This method of fracture healing is characterised by small movements at the fracture site, but too much motion can delay or even prevent healing. In contrast, rigid fixation, which restricts movement entirely, can delay this process. Secondary healing typically occurs in nonoperative fractures (fractures treated without surgery) and in cases where minimal movement is present, such as with external fixation, internal fixation of complex comminuted fractures (fractures with multiple fragments), or intramedullary nailing (insertion of a metal rod inside the bone) (Marsell & Einhorn, 2011).

In intramembranous ossification, new bone forms directly without first creating cartilage. However, endochondral ossification follows a different path, where cartilage is initially formed and later replaced by bone. Secondary fracture healing progresses through three overlapping phases: the inflammatory phase, the repair phase (fibroplasia), and the remodelling phase. These stages are not entirely separate, as some processes continue into the next phase while new processes may start in the previous phase (Tsiridis et al., 2007). The healing process begins with a hematoma (a collection of blood) forming at the fracture site, triggering an inflammatory response. In the next phase, angiogenesis (the formation of new blood vessels) occurs, and granulation tissue develops, followed by cartilage formation, creating the callus. As healing continues, the cartilage hardens through cartilage calcification (endochondral ossification) and is gradually replaced by bone (Oryan, Ahmad et al., 2015).

Finally, the remodelling phase restores the original structure and strength of the bone, a process that can take several months to years (Dimitriou et al., 2005; Marsell & Einhorn, 2011; Oryan, Ahmad et al., 2015; Tsiridis et al., 2007)

2.2.3 Stages of bone healing

2.2.3.1 Inflammatory Phase

The first stage of fracture healing begins immediately after injury, and this process involves blood clotting (haematoma formation) and inflammation. It starts within 12 to 14 hours of the trauma, as damage to the blood vessel acting as a trigger, which leads to platelet aggregation (clustering of blood cells), blood coagulation (clotting), and clot formation at the fracture site. The resulting blood clot serves as a foundation for the arrival of inflammatory cells, endothelial cells (cells lining blood vessels), and fibroblasts (cells responsible for tissue repair) (LaStayo et al., 2003). The acute inflammatory response peaks within the first 24 hours and generally concludes by 7 days post-injury. The first immune cells to reach the site are neutrophils (white blood cells that fight infection). They are soon replaced by macrophages, lymphocytes, and plasma cells, which assist in tissue repair. Macrophages not only clear out damaged tissues and debris through a process called phagocytosis but also release growth factors and cytokines (proteins that regulate cell signalling and healing) (Dimitriou et al., 2005).

Key growth factors released by platelets, stimulates mesenchymal stem cells (cells capable of developing into different tissue types), which likely originate from the periosteum (bone covering), bone marrow, bloodstream, and nearby soft tissues. These stem cells transform into specialised fibroblasts (connective tissue cells), angioblasts (blood vessel-forming cells), chondroblasts (cartilage-forming cells), and osteoblasts (bone-forming cells) to aid tissue regeneration. In addition, mechanical loading (weight-bearing activities and movement) acts as a stimulant for these stem cells to become fibroblasts, chondroblasts, and osteoblasts, to promote further healing. The growth factors initiate granulation tissue formation, which leads to the creation of a soft callus. This stage is followed by the formation of a hard callus, marking the second phase of bone healing. At this point, the bone remains weak, with low stiffness and tensile strength, making it vulnerable to excessive forces or weight-bearing stress (Dimitriou et al., 2005; LaStayo et al., 2003).

2.2.3.2 Repair or regenerative phase

In this section, the regenerative phase of bone fracture healing is explained.

2.2.3.2.1 Fibrocartilage (soft callus) formation

Following haematoma formation where the blood clots at the injury site, a connective tissue called granulation tissue is formed during early healing process. This tissue is rich in Fibrin, a

protein that helps clotting. This forms a natural scaffold – a temporary structure that supports cell growth and tissue repair. Within this scaffold, a process called endochondral ossification (bone formation through cartilage mineralisation) takes place, primarily between the fractured bone ends and around the outer layer of bone. Since these regions are less stable, cartilaginous tissue (soft cartilage) develops into a soft callus, which provides initial fracture stability (Dimitriou et al., 2005; LaStayo et al., 2003).

The cartilaginous callus undergoes mineralisation, a process where calcium and other minerals harden the tissue. Over time, this cartilage is reabsorbed and replaced with bone, marking a significant phase in fracture healing. Initially, this semi-rigid callus is lacking blood vessels, but as it transitions into woven bone (newly formed, unorganised bone), vascular invasion occurs, meaning new blood vessels grow into the area to supply oxygen and nutrients. Cartilage-forming cells, known as chondrocytes, come from stem cells that have the potential to develop into different types of tissue. These chondrocytes multiply quickly and create a gel-like substance called the cartilaginous matrix, which gives cartilage its strength and flexibility. An adequate blood supply is essential for effective bone healing. In endochondral ossification, chondrocytes eventually die, a term known as Apoptosis, and the cartilage matrix breaks down. This allows for new capillaries (tiny blood vessels) to grow into the area (Dimitriou et al., 2005; LaStayo et al., 2003).

This angiogenesis (formation of new blood vessels) is stimulated by some factors with which the VEGF (Vascular Endothelial Growth Factor), is particularly crucial for restoring blood flow to the fractured bone. In laboratory studies on mice, rats, and rabbits, soft callus formation peaks between 7 to 9 days after injury, signifying a critical period in the early healing process (Dimitriou et al., 2005; LaStayo et al., 2003).

2.2.3.2.2 Hard callous formation

As the hard callus develops and the calcified cartilage is replaced by woven bone, the callus becomes stronger and more stable. The calcified cartilage helps trigger the formation of new blood vessels (angiogenesis) in the regenerating tissue and attracts osteoclasts and osteoblasts, which are the cells responsible for breaking down old cartilage and forming new bone to the fracture site. Osteoclasts clear away the cartilage matrix, making room for the woven bone to take its place. A crucial step in fracture healing is the formation of a cartilaginous callus, which is gradually hardened, broken down, and replaced with bone (Oryan, Ahmad et al., 2015).

2.2.3.3 Bone remodelling phase

In this the remodelling phase of bone fracture healing is explained.

2.2.3.3.1 Differentiation of the woven bone

The third stage of fracture healing is bone remodelling, where the newly formed woven bone is gradually replaced with lamellar bone, a more organised and structured type of bone. Lamellar bone can be either cortical bone (dense, outer bone layer) or trabecular bone (spongy, inner bone layer) (Dimitriou et al., 2005). This phase can take several months to years and aims to restore the bone's original structure, strength, and function. During this stage, the hard callus is slowly reshaped to form the bone cortex (outer shell) and medulla (inner cavity). The process is supported by neovascularisation, meaning the formation of new blood vessels continues to ensure proper nutrient supply to the healing bone. A key component of this phase is the activity of osteoclasts, specialised bone cells that break down and absorb calcified bone tissue. This breakdown clears space for osteoblasts, which are cells responsible for producing osteoid; the unmineralised part of new bone (Marsell & Einhorn, 2011).

Over time, this osteoid becomes mineralised, strengthening the fracture site. Several biochemical signals in the blood cells play a role in osteoclast formation, ensuring the bone is properly remodelled (Dimitriou et al., 2005; Marsell & Einhorn, 2011).

2.2.3.3.2 Alignment and maturation of the newly formed tissue

At this stage of bone healing, osteoclasts begin to break down and reshape the newly formed bone tissue in the injured area. This process ensures that the repaired bone matches the structure of the original, uninjured bone. At the same time, osteoblasts continue to deposit osteoid. They also add calcium phosphate, which strengthens and hardens the bone matrix (Marsell & Einhorn, 2011).

As a result of osteoclast activity, the bone's outer width decreases, making it more similar in shape to the original bone. However, the internal bone structure becomes denser, improving its strength. As this phase progresses, the number of active cells involved in bone healing gradually decreases, and bone density continues to improve (Oryan et al. (2015)(Marsell & Einhorn, 2011). This occurs naturally as the individual resumes weight-bearing activities and movement, which further stimulates bone remodelling and enhances its strength (Marsell & Einhorn, 2011; Oryan, Ahmad et al., 2015) .

2.3 Sprain injury

A sprain on the other hand simply relates to a damaged ligament. Most damaged ligaments are caused by injuries that forces the joint to beyond its range of motion. Sprain occurs mostly where two or more bones are connected, with majority of sprain injuries occurring in the ankle, wrist and knee. Sprains are mild when compared to fractures with symptoms including minor swelling and bruises (Baoge et al., 2012). When there is an incidence of a sprain injury, the symptoms often mimic those of a fracture, including swelling and raised temperature at the affected site. The injury and healing process is divided into three distinct stages, similar to those seen in fractures:

- Destruction and Inflammation Stage (1-3 days): The initial response, where inflammation occurs to remove damaged cells.
- Repair Phase (3-4 weeks): The body begins repairing the damaged ligament and forming temporary connective tissue.
- Remodelling Phase (3-6 months): The tissue strengthens and reorganises to restore function (Järvinen et al., 2005).

Unlike fracture healing, where a callus forms during repair, sprain healing involves the regeneration of muscle fibres (myofibres) and the formation of scar tissue (fibrinous tissue), which is later restructured in the remodelling phase (Baoge et al., 2012; Järvinen et al., 2005).

2.4 Common imaging techniques for fracture diagnosis

In this section the common bone fraction diagnosis modalities are explained.

2.4.1 Conventional radiography

X-ray radiography has long been the standard and most widely used imaging technique for diagnosing and monitoring bone fractures. It operates by passing a controlled dose of ionising radiation through the body, which is absorbed at different rates by various tissues. Dense materials like bone absorb more X-rays and appear white on the resulting image, while softer tissues appear darker. This contrast enables clinicians to detect breaks, cracks, or misalignments in bones. Traditionally, X-rays have also been employed to evaluate the healing process of fractures. One of the key signs of bone recovery that clinicians look for is the presence of callus formation, a sign that new bone is forming at the fracture site (Firoozabadi et al., 2008).

However, the interpretation of X-ray images is somewhat subjective, and questions remain about their accuracy in measuring bone strength and complete healing. Ultimately, while radiographs

are useful, they may not fully reflect the functional strength of the healed bone (Firoozabadi et al., 2008).

2.4.2 Computed tomography

Computed Tomography (CT) is an advanced imaging technique that relies on X-ray technology to produce detailed cross-sectional images of the body's internal structures. Unlike standard X-rays that provide two-dimensional views, CT scans capture multiple thin slices of tissue, allowing for more accurate visualisation of organs, bones, and other internal features. The CT system operates by rotating a powerful X-ray tube and a high-resolution digital detector array around the patient. As the machine rotates, it collects numerous data points from different angles, and these individual measurements are known as projections. Specialised computer algorithms then process these projections through a technique called tomographic reconstruction to generate detailed cross-sectional images. Modern CT technology enables the rapid acquisition of multiple adjacent slices, which can be digitally assembled into three-dimensional (3D) models for enhanced diagnostic insight (Firoozabadi et al., 2008; Relyea-Chew & Chew, 2019).

Additionally, advanced CT methods such as Quantitative Computed Tomography (QCT) and Volumetric Computed Tomography (VCT) offer improved capabilities in assessing bone density and volume, further enhancing clinical decision-making (Firoozabadi et al., 2008). A cross section of the CT showing different aspect of the wrist is shown in Figure 2.4 (Relyea-Chew & Chew, 2019).

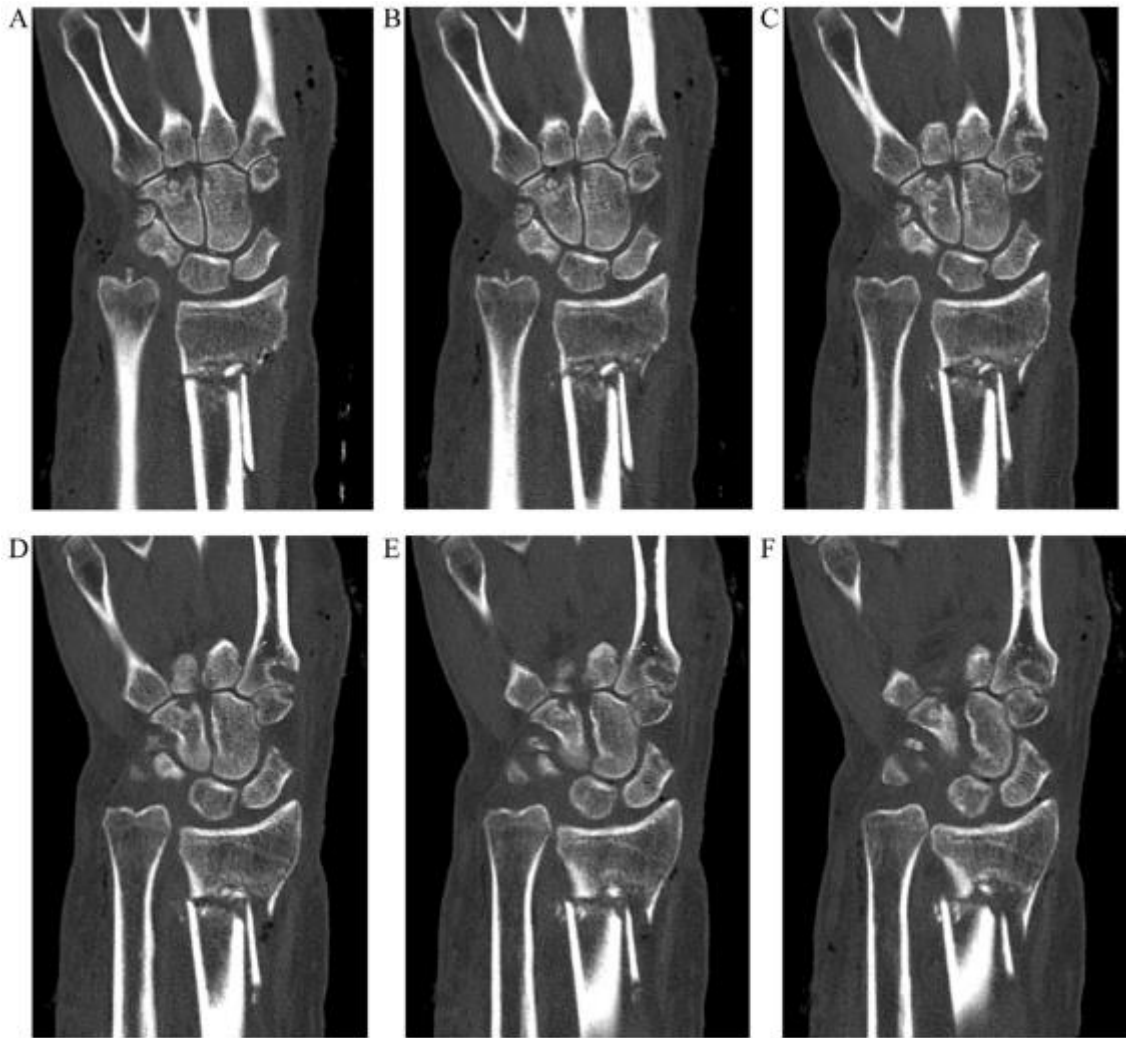


Figure 2. 4. Dorsal to volar (A – F) CT scan of a fractured wrist area (Relyea-Chew & Chew, 2019).

2.4.3 Magnetic resonance imaging

Magnetic Resonance (MR) microscopy which comprises high-resolution MR (hrMRI) and micro-MR imaging, has been used extensively in both research and clinical settings, particularly for assessing the microarchitecture of trabecular bone in conditions related to bone fragility, such as osteoporosis. MR imaging is a sophisticated technology that operates using strong magnetic fields and radiofrequency (RF) pulses. These RF waves excite hydrogen protons within the body's tissues, and the emitted signals are captured and processed to generate detailed images. One of the major advantages of MR imaging is its non-invasive nature and the absence of ionising radiation, making it a safer alternative to other imaging techniques like X-rays or CT scans. MR can produce three-dimensional (3D) images in any plane or orientation, offering exceptional versatility. In bone imaging, trabecular structures appear as dark or negative spaces due to the

high signal intensity emitted by surrounding bone marrow, which contains a large concentration of fat and water (Firoozabadi et al., 2008).

Despite its advanced capabilities, MR imaging comes with certain limitations. The process of acquiring MR data tends to be more expensive and time-consuming compared to CT imaging. Additionally, the interpretation of MR images is more technically demanding due to the complexity of signal behaviour and image processing techniques. While MR holds significant promise for quantitatively analysing trabecular bone structure in both living subjects (*in vivo*) and specimens (*in vitro*), its widespread application in fracture healing studies remains limited. This is largely due to the higher operational costs and complexity of MR systems compared to micro-CT (mCT), which is more frequently used in experimental research on bone repair and regeneration (Firoozabadi et al., 2008).

2.4.4 Ultrasonography

In this section, the use of ultrasonography for bone fracture detection is explained.

2.4.4.1 Ultrasound origin and development.

The use of ultrasound in medical diagnostics dates back to the 1940s, marking a pivotal advancement in non-invasive imaging technology. The earliest clinical application employed A-mode (amplitude mode) ultrasound, which provided one-dimensional images based on echo amplitude. This was followed by the development and sequential integration of more sophisticated imaging modes, including B-mode (brightness mode), M-mode (motion mode), and Doppler mode (D-mode), each offering improved visualisation and diagnostic capabilities (Kalogiouri & Samanidou, 2021; Shinohara et al., 2021). Ultrasound quickly gained popularity among healthcare professionals and patients alike due to its non-invasive nature, ease of use, real-time imaging capability, and affordability compared to other diagnostic imaging modalities such as CT or MRI. Its ability to be safely repeated without ionising radiation further enhanced its clinical utility across various patient populations.

A significant milestone in the evolution of ultrasonic diagnostics was the introduction of colour Doppler imaging. This advancement allowed for dynamic visualisation of blood flow within the human body, enhancing the assessment of vascular structures, cardiac chambers, the biliary system, and other soft tissue organs (Shinohara et al., 2021). By enabling clearer, more intuitive sonographic representations, colour Doppler greatly improved the clinician's ability to evaluate both anatomical and functional characteristics of internal structures. Figure 2.5 (Hareendranathan

et al., 2023) shows an example of ultrasound (A), viz-a-viz the X-ray image (B) of a patient with wrist fracture.

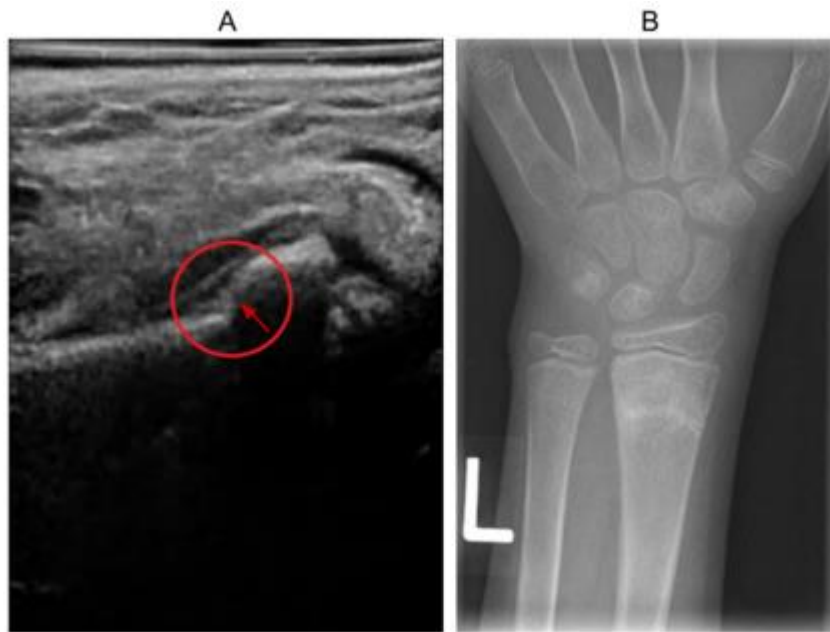


Figure 2. 5. Showing US (A) and X-ray (B) images of a patient with wrist fracture. The fracture site is denoted by the red arrow in the US image (Hareendranathan et al., 2023).

The field of ultrasound has seen remarkable growth in the last decade, which is driven by technological innovations in imaging equipment and transducer design. Modern ultrasound systems now include high-resolution probes that can be used in body cavities or intraoperatively, significantly broadening the scope of diagnostic and therapeutic applications. These developments have facilitated early and accurate detection of pathological conditions, even in complex clinical scenarios (Fritz et al., 2022). High-frequency ultrasound, typically ranging from 3 MHz to 30 MHz, is now routinely used in clinical settings. Higher frequencies provide superior spatial resolution, allowing for detailed visualisation of superficial structures and subtle tissue changes, which is particularly beneficial in dermatology, musculoskeletal imaging, and ophthalmology. Another critical advancement is grey-scale ultrasound imaging. This technique leverages the amplitude of returning echoes to create varying shades of black and white on the sonogram, enabling the differentiation of tissue types based on echogenicity. Strong echoes result in brighter (white) areas, while weaker echoes produce darker (grey to black) regions (Yi, J. et al., 2022) . This method is highly effective in evaluating soft tissue masses, distinguishing cystic from solid lesions, and aiding in the diagnosis of conditions such as placenta accreta in postpartum patients (Yi, J. et al., 2022).

2.4.4.2 Application of US in fracture diagnosis and treatment

As discussed in the previous sections, bone fractures occur when direct or indirect trauma disrupts the structural integrity of bone tissue, particularly the *trabeculae*, which are the small, lattice-like structures inside bones that provide strength and support. When these internal bone structures are damaged, it can result in displacement (where the bone parts shift out of their normal alignment), dislocation (where the bone is pushed out of its joint), and damage to nearby soft tissues, such as muscles, ligaments, and tendons. In healthy bones, the outer surface reflects ultrasound waves very strongly, creating a bright interface that can prevent deeper areas from being visualised effectively. This is due to the dense nature of bone, which blocks ultrasound penetration beyond the bone surface (Auten et al., 2019). However, if the bone surface (known as the *periosteum*, a thin layer of connective tissue covering the bone) is damaged, whether due to thinning, fracture, or disease, ultrasound can become a valuable tool for visualising abnormalities beneath this layer, such as *subperiosteal lesions* (injuries or bleeding occurring under the periosteum) (Du et al., 2022) .

High-frequency ultrasound, particularly using probes in the 8–12 MHz range, is commonly employed to examine the outer surface of bones in areas where patients report pain or swelling. These high-frequency probes offer better image resolution, which allows clinicians to clearly observe both soft tissue swelling and the continuity of the bone cortex (the hard outer layer of the bone). When a fracture is suspected, ultrasound can identify breaks or interruptions in the bone surface (Du et al., 2022).

Clinicians can then use a multi-angle scanning approach to examine the area thoroughly, often revealing stepped or jagged elevations in the bone surface, which are signs of displaced bone fragments. Research conducted in studies by (Masoudi & Naraghi, 2022; Natrajan et al., 2021) have indicated that certain fractures, particularly those associated with soft tissue swelling or tiny bone fragment separation (*avulsion fractures*), may not be visible on X-ray but can be detected with ultrasound.

Ultrasound can also be used for the diagnosis of fractures of the radius and ulna. Fractures of the ulna and radius are especially common in children aged 5 to 12. At this stage of development, the *epiphyseal line* (the growth plate region in long bones) is relatively weak, making the bones more susceptible to fractures. Traditional X-ray imaging requires the child to move the affected limb into various positions to capture clear images, which can cause additional pain. In contrast, ultrasound can be performed with minimal limb movement, significantly reducing discomfort

and eliminating exposure to radiation. The main ultrasound indicators of ulnar or radial fractures include disruptions in the normally continuous, bright echo of the bone cortex, abnormal bending or deformation of the bone, accumulation of blood under the periosteum (*subperiosteal hematoma*), and swelling of surrounding soft tissues (*edema*) (Auten et al., 2019).

A study in the Netherlands (Epema et al., 2019) comparing ultrasound with X-ray for diagnosing distal forearm fractures in children aged 0 – 14 found that ultrasound achieved a 92% diagnostic accuracy when compared to X-rays with the added benefits of being faster and less painful. Another study by (Auten et al., 2019), showed that ultrasound had a sensitivity of 100% and specificity of 92–93% in identifying inadequate fracture reduction (improper realignment) in children with ulnar fractures. This underscores ultrasound's reliability as an effective and child-friendly method for diagnosing and evaluating forearm fractures.

2.4.4.3 Application of US for fracture treatment

Ultrasound (US) plays a growing role in the treatment and monitoring of bone fractures due to its safety, non-invasiveness, and ability to be repeated without radiation. It is widely used for fracture localisation, guiding surgical procedures, anaesthesia, and monitoring healing. Preoperative US can improve fracture end alignment, reduce incision size, and shorten surgical time, especially in patients with rib or hip fractures. It also helps visualise surrounding nerves and blood vessels, aiding complex fracture reductions and minimizing damage to vital structures (Auten et al., 2019).

In anaesthesia, US-guided nerve blocks such as lumbar or brachial plexus blocks, offers effective pain control with fewer complications, particularly in elderly patients. For accurate fracture localisation, US has shown strong results, comparable to CT scans, in nasal and rib fractures. US also effectively monitors healing stages discussed in earlier sections. It detects early signs of callus formation and bone remodelling earlier than X-rays. Advanced techniques, like 3D ultrasound reconstruction, further enhance fracture assessment. Additionally, low-intensity pulsed ultrasound has been found to accelerate bone healing by stimulating cellular activity in the fracture site (Du et al., 2022).

2.4.4.4 Advantages and limitations of ultrasound applications

High-frequency ultrasound is a powerful diagnostic imaging tool that operates at frequencies typically between 7-15 MHz and presents itself as a powerful tool especially in the effective detection of small or subtle bone fractures, often referred to as occult fractures that may not be

immediately visible on standard X-ray images. Because of its enhanced sensitivity, it has proven to be particularly valuable in identifying fracture lines that would otherwise go unnoticed using traditional radiography (Auten et al., 2019; Du et al., 2022). One of the most important advantages of ultrasound is its ability to image from multiple angles and directions, allowing clinicians to examine the affected area thoroughly without subjecting patients to repeated exposure to ionising radiation. This makes ultrasound a safer and more suitable option for vulnerable populations, such as children, pregnant women, and patients who require frequent imaging (Du et al., 2022). Despite its many benefits, ultrasound is not without limitations. Its effectiveness is influenced by the physical properties of the tissues being scanned. For example, ultrasound waves cannot penetrate bone tissue effectively, which means it cannot provide detailed images of internal bone structures. Instead, it focuses mainly on the bone surface and surrounding soft tissues. This makes it less effective for complex fractures that involve deeper bone layers. Additionally, the field of view in ultrasound imaging is relatively limited compared to other modalities such as CT or MRI. Moreover, ultrasound diagnostics can be highly operator-dependent, meaning the quality and accuracy of the results are significantly influenced by the skill, experience, and subjective interpretation of the clinician performing the scan (Auten et al., 2019; Du et al., 2022; Epema et al., 2019).

Given these limitations, ultrasound is often best used in conjunction with other imaging techniques when dealing with complicated or unclear cases. For example, if ultrasound results are inconclusive due to poor image quality or anatomical complexity, further imaging with X-ray, CT, or MRI may be necessary to achieve a definitive diagnosis. CT scans provide detailed cross-sectional images of bone structures, while MRI is particularly effective at visualising soft tissue and bone marrow changes associated with certain types of fractures (Du et al., 2022)

In conclusion, high-frequency ultrasound offers a valuable, safe, and radiation-free method for detecting fractures, particularly in sensitive or early-stage cases. However, it should be viewed as part of a comprehensive diagnostic toolkit rather than a stand-alone solution, especially when image clarity or fracture complexity presents diagnostic challenges.

2.4.5 Thermography as a diagnostic tool.

In this section the use of thermography as a diagnostic tool were presented.

2.4.5.1 Thermography origin and development

Thermography can be defined as a scanning method that produces an image-like map of temperature variations across the body's surface. Because human physiology tightly regulates

core and tissue temperatures within a narrow range under controlled environmental conditions, departures from that range may signal disease processes, structural abnormalities, or functional defects (Winsor & Winsor, 1985). Fever has long been a clinician's most visible clue to disease. Early researchers such as Hippocrates described placing damp clay on the skin and watching how quickly it dried over inflamed swellings, which implies an early practical recognition that abnormal heat accompanies illness. Objective temperature measurement came much later. In the late sixteenth century, Galileo devised a thermoscope, a glass-tube device that registered temperature change but being open to the air, also responded to barometric pressure (Ring & Ammer, 2012).

Thermographic technique relies on the infrared radiation naturally emitted by skin; an infrared detector, tuned to specific wavelengths, converts these emissions into measurable signals that represent local heat patterns (Lahiri et al., 2012). In this way, thermographic images offer a visual record of relative temperature differences rather than absolute anatomical detail. During the 1970s, this approach was viewed as a promising medical imaging modality, yet its clinical momentum diminished when early instruments lacked maturity and misconceptions about its capabilities spread. As a result, adoption receded despite the method's theoretical appeal and physiological basis in routine clinical practice then (Kesztyüs et al., 2022; Winsor & Winsor, 1985).

2.4.5.2 Infrared thermography

Infrared thermography (IRT) for non-destructive testing and evaluation is used to reveal subsurface features such as defects, voids, or other anomalies usually by tracking temperature contrasts (ΔT) that appear on the surface while the area is observed with an infrared camera. Any object warmer than absolute zero ($-273\text{ }^{\circ}\text{C}$, 0 K) including the human skin and many engineering materials emits electromagnetic radiation which travels at the speed of light; continuously radiate energy that can be sensed remotely (Avdelidis et al., 2011; Balaras & Argiriou, 2002). This thermal emission originates from the random motion of particles and charge carriers within the material; their agitation produces fluctuating electromagnetic fields, emitted as photons. The wavelength and frequency of the particles can be measured since the photon particles moves at the speed of light, using the formula (Arthur et al., 2011).

$$c = \lambda f \quad 2.1$$

where $C = \text{Speed of light}$, $\lambda = \text{Wave length}$ and $f = \text{frequency}$

IRT exploits this physics by detecting radiation within defined spectral bands, most commonly the mid-wave infrared (about 2-5.6 μm) and the long-wave infrared (about 8-14 μm) as shown in Figure 2.6 (Plastropoulos et al., 2023) and converting the measured intensity into an apparent temperature map). These two spectral bands are commonly used because of their low atmosphere absorption (Plastropoulos et al., 2023). During natural cooling/heating or under external stimulation, in-depth inhomogeneities disturb heat flow, generating surface ΔT patterns that portray underlying problems. Because it requires no contact and does not damage the material being measured, IRT enables rapid surveys, wide-area coverage, and repeated monitoring.

The radiant heat emitted per unit area from a surface is proportional to its emissivity and to the fourth power of its absolute temperature. This relationship is given by the Stefan–Boltzmann law (Balaras & Argiriou, 2002):

$$q'' = \sigma \varepsilon T^4 \quad 2.2$$

Where q'' is the hemispherical total emissive power (radiated energy per unit area, $\frac{\text{W}}{\text{m}^2}$)

σ is Stefan-Boltzmann constant ($5.67 \times 10^{-8} \frac{\text{W}}{\text{m}^2} \text{K}$)

ε is total hemispherical emissivity of the surface ($0 < \varepsilon < 1$)

T is the surface temperature in Kelvin (K)`

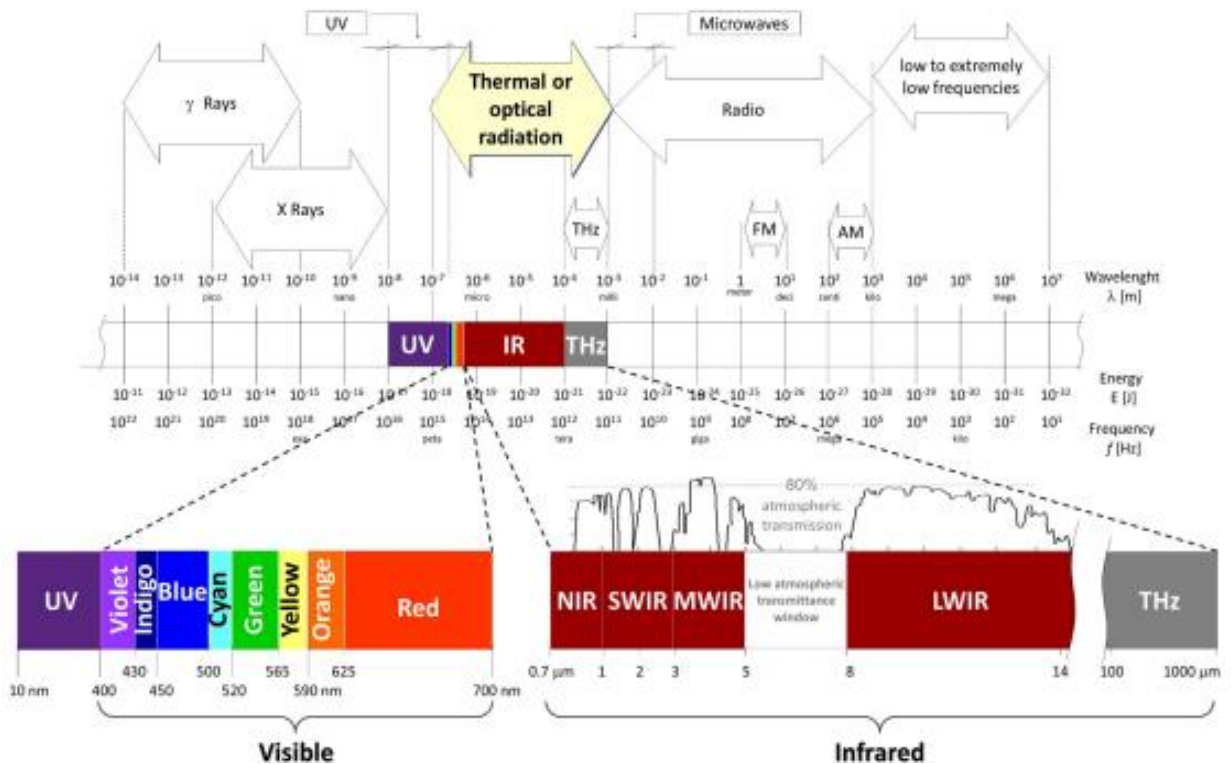


Figure 2. 6. Electromagnetic Spectrum showing the IR band (Plastropoulos et al., 2023).

In IRT, the main challenge is emissivity which can be regarded as the efficiency with which a surface emits thermal radiation. An IR camera does not measure temperature directly; it senses radiant energy from the surface and converts that signal to an apparent temperature using an assumed emissivity value. Emissivity (ϵ) is a surface property defined as the ratio of a real surface's emission to that of an ideal blackbody at the same temperature, ranging from 0 (perfect reflector, mirror-like) to 1 (perfect emitter, blackbody) (Avdelidis & Almond, 2004).

Crucially, ϵ depends on wavelength, temperature, and surface condition (roughness, oxidation, coatings, and contamination), so incorrect settings can produce large temperature errors.

Polished metals such as aluminium or steel typically have low ϵ and high reflectance, acting like infrared mirrors that reflect ambient and operator heat into the camera and distort readings (Avdelidis et al., 2011). To address this, practitioners raise the effective emissivity of the test spot: applying a thin high-emissivity coating (e.g., matte black paint or water-based black coatings), using high- ϵ adhesive tape or paint dots, lightly roughening the surface, or relying on an oxide layer when appropriate. These steps, combined with proper camera configuration (matching ϵ to the camera's spectral band, often 3-5 μm or 8-14 μm) and, if needed, reflected-temperature correction or a blackbody reference, enable reliable thermographic surveys

(Avdelidis & Almond, 2004; Plastropoulos et al., 2023). Examples of IRT images are shown in the Figure 2.7 (Kastberger & Stachl, 2003; Owen, R. et al., 2017; Reed et al., 2020):

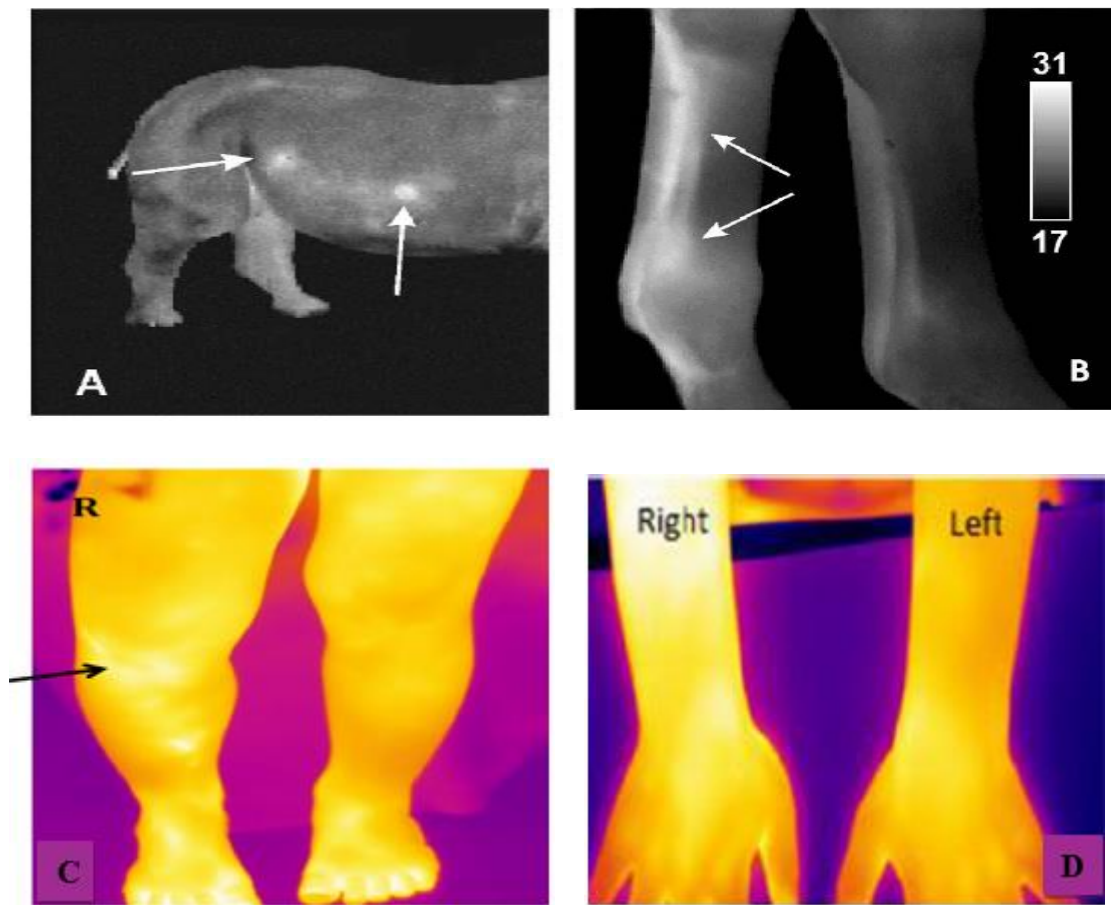


Figure 2. 7. A. IR thermography helping in the diagnosis of pregnancy in a rhinoceros (Kastberger & Stachl, 2003), B. IR thermographic image showing inflamed hindleg patellar damage in a horse (Kastberger & Stachl, 2003) C. IR thermal image helping to diagnose fracture in the right tibia of a toddler (Owen, R. et al., 2017) D. IR thermal image helping to detect wrist fracture in children (Reed et al., 2020).

2.4.5.3 Approaches to deploying IRT

IRT is typically applied in two principal modes: passive and active (Avdelidis et al., 2011).

2.4.5.3.1 Active thermography

Active thermography, by contrast, introduces an external stimulus to generate diagnostically useful thermal contrasts. Common excitations include optical flash lamps, continuous heat lamps, halogen sources, hot- or cold-air guns, electrical heating, ultrasound, or induction. The applied thermal pulse or steady heating creates transient or steady-state temperature differences that

diffuse from subsurface features to the surface, enhancing visibility of defects and enabling depth estimation (Avdelidis et al., 2011). The active approach set up is pictured in Figure 2.8 (Avdelidis et al., 2011).

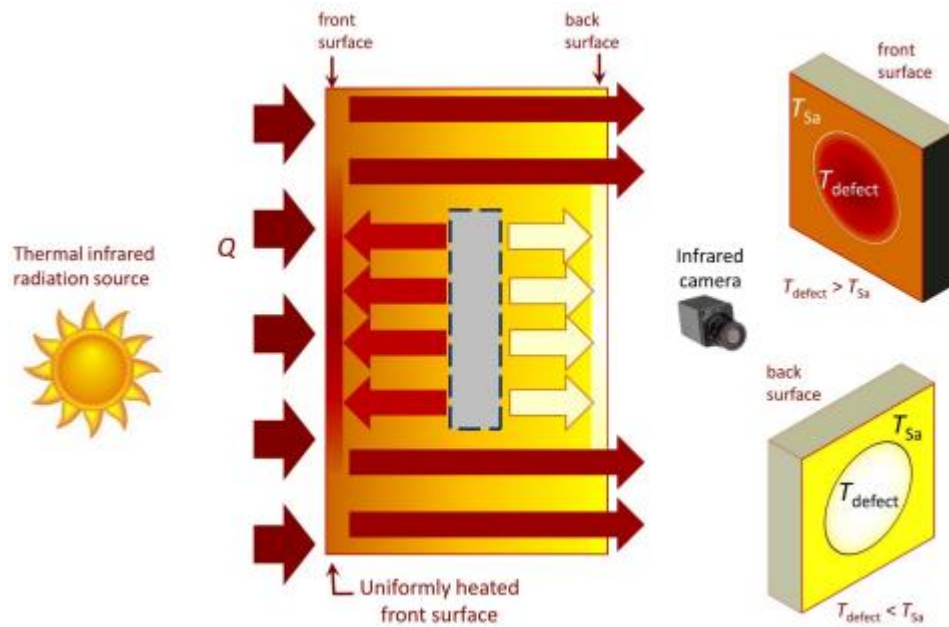


Figure 2. 8. Active thermography set up (Avdelidis et al., 2011).

2.4.5.3.2 Passive thermography

In passive thermography, the target naturally sits at a temperature different from its surroundings (often warmer), so the camera observes existing heat patterns without intervention. This approach is widely used for medical applications that look for asymmetric or abnormal heat signatures, building and infrastructure surveys (e.g., heat loss, moisture ingress), maintenance of electrical and mechanical equipment (overheating or unbalanced loads), process monitoring. Passive inspections are most often qualitative, which implies that they are mostly used in highlighting irregularities, hot spots, or cold bridges. With careful calibration and emissivity control, it can however be used for quantitative measurements (Avdelidis et al., 2011). The passive approach is pictured in Figure 2.9 (Avdelidis et al., 2011).

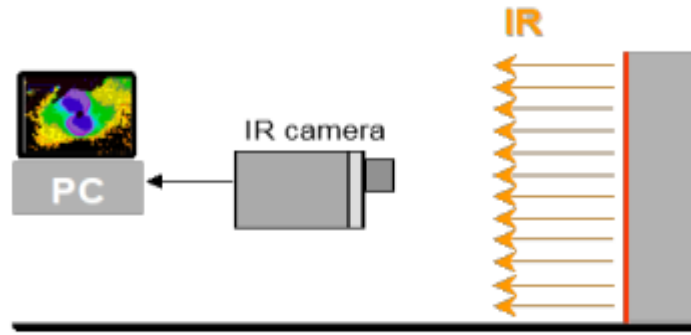


Figure 2. 9. Passive Thermography (Avdelidis et al., 2011).

2.4.5.4 IRT for medical image diagnosis

Several studies have used thermal images to diagnose one form of disease or the other. In an earlier pilot study (Reed et al., 2020) used infrared thermal imaging to differentiate between fracture and sprain in an experiment with children in a local hospital. Temperature of surfaces referred to as ROI were measured and compared with the temperature of other regions of the body. Statistical tools such as t-test were used to quantify the mean temperature difference. Result showed that there was a significant temperature difference with higher values in fracture than in sprain for the selected ROI. In another study (Owen, R. et al., 2017) used thermal imaging to diagnose acute undifferentiated limb in 30 children recruited at a local hospital. Also, statistical tools were also used to evaluate the recorded temperature. In this case the median temperature difference was considered due to the skewness of the mean temperature as a result of small sample size. Their results also showed significant temperature difference in different areas of the body such as the hip and thighs, knee and lower leg. Eleven patients were recruited in the study (Fane De Salis et al., 2018) that used IR thermal imaging to determine the effects and extent of vertebrae fractures in patients with Osteogenesis Imperfecta (OI). These studies indicated there was a sufficient mean temperature increase in the ROI of the subjects compared to the skin of healthy bones using images from the thermal camera.

2.4.5.4.1 Artificial intelligence based diagnosis using IRT.

Once there is a sufficient temperature difference between the ROI as compared with the related reference region, AI classification algorithms can therefore be used to provide a computer-aided screening of the focused body part (Hossein Ghayoumi zadeh et al., 2013; Snehalatha & Sangamithirai, 2021) . (Snehalatha & Sangamithirai, 2021) developed a custom deep learning neural network (DNN) algorithm to classify thermal images with the abdomen, forearm, and shank region as the ROI. They recruited 50 obese and 50 healthy patients. CNN-2 algorithm was used and an accuracy of 92% was obtained.(Hossein Ghayoumi zadeh et al., 2013) developed a

non-clinical approach to diagnose breast cancer using IRT imaging and a combination of ANN and genetic Algorithm (GA). IRT images of 200 subjects were obtained for the study. Their system had an accuracy of 70%. (Vardasca et al., 2019) compared different AI techniques to classify IRT imaging of 39 patients with Diabetic Foot Ulcers (DFU). The processed images were passed through 3 decision support machine learning algorithms namely: ANN, Support Vector Machine (SVM) and k-Nearest Neighbour (k-NN), for classification of the images. From the comparison of their AI techniques, they found that k-NN of 5 neighbours provided the best classification of the disease with an accuracy of 81.25%. These findings shows that AI algorithms can be an alternative diagnosis tool when applied to IRT images. Although there are not lot of work that have used IRT imaging especially for bone fracture, the literature review in chapter 3 provides a comprehensive view of AI techniques using different image modalities and disease analysis.

2.5 Summary

In this chapter surveys bone biology, fracture healing, and imaging was presented, providing context for infrared thermography (IRT) and AI-aided diagnosis. It explains bone composition (water, organic collagen matrix, inorganic hydroxyapatite), architecture (cortical vs trabecular), and key cells (osteoblasts, osteoclasts, osteocytes) that drive growth, remodelling, and repair. The difference between fracture and sprain is also presented. It then reviews diagnostics: radiography for detection and monitoring, MRI for radiation-free microarchitecture and ultrasonography. Finally, it introduces IRT's physics (Stefan–Boltzmann, emissivity), passive vs active deployment, and practical issues (emissivity control, calibration). Studies showing how thermal features can help distinguish fractures from sprains, and machine-learning models (e.g., SVM, k-NN, CNNs) can classify IRT images were also presented, highlighting IRT's potential as a non-contact adjunct to existing pathways.

Chapter 3 – Literature Review

In this chapter an overview of artificial intelligence methods and studies for medical image processing are presented.

3.1 Introduction

The use of imaging for diagnosis in healthcare is substantial, amounting to about 100 billion dollars globally per year (Jeong et al., 2022). Mounting pressures on the healthcare facilities and the market for imaging diagnosis have led to increasing demands for diagnostic excellence in the clinical setting due to rising number of clinical images, image complexities and faster results, as demanded by clinicians. As a result, the need for new technologies is centred around providing solutions that will increase the effectiveness of the clinical process, improve the healthcare systems and provide accurate diagnosis for the patients, while improving care quality. Therefore, there have been high demands for technologies that can aid the automation of workflows associated with the use of medical imaging for diagnosis, leading to advances in the use of artificial intelligence (AI) methods such as deep learning to assist radiologists in analysing complex image datasets (Yu-Jen Chen et al., 2015).

3.1.1 Deep neural network overview

Deep learning is a subfield of machine learning which leverages the artificial neural networks (ANN) architecture to acquire knowledge from large datasets and perform intricate operations. One of the main advantages of deep learning techniques is their ability to mimic the information processing complexity of the human brain (Yu-Jen Chen et al., 2015; Zhu, Zirui, 2024). The field has been a studied since the 1980s but gained prominence in recent years. This is because of access to large datasets for model training, improved algorithm development and increased processing power of microprocessors. The structure of an ANN is an interconnection of nodes (sometimes referred to as processing elements or neurons) which can span up to multiple layers, depending on the intricacy of the tasks and capabilities of hardware resources. Each node gathers information from the previous layer and then transmits to the subsequent layer based on the configured characteristics and set parameters. The values of the parameters used are normally initially randomised but then iteratively updated during training based on a set learning rate (Sarmadi et al., 2024). A deep learning network extracts deeper and intricate information as its number of architectural layers increases, resulting in optimised performance with large datasets

and training iterations. This in turn ensures precise recognition of patterns in the data (Zhu, Zirui, 2024).

3.1.2 Deep learning in medical imaging, classification and segmentation

ANN and DL has seen extensive applications in medical image analysis. It has been used for different medical image modalities including X-ray radiographs, Computed Tomography (CT), Ultrasound (US) and Magnetic Resonance Imaging (MRI) scans to provide predictive diagnosis and treatment. Subtle and intricate patterns presented by these medical images are effectively identified by the adapted deep learning approach, thereby providing a means to automate the feature extraction process. Image feature extraction or selection is a process that can be performed manually by a qualified specialist, but this can be time consuming and subjective. When deep learning models are properly trained, they have an ability to accurately identify lesion or tumours, examine membranes and tissues for differences and do diverse medical related tasks thereby providing accelerated diagnostic outcomes. Therefore, deep learning is emerging into the realm of medical images analysis assisting with diagnosis. A summary of deep learning applications is shown in Figure 3.1 (Shobayo & Saatchi, 2025)

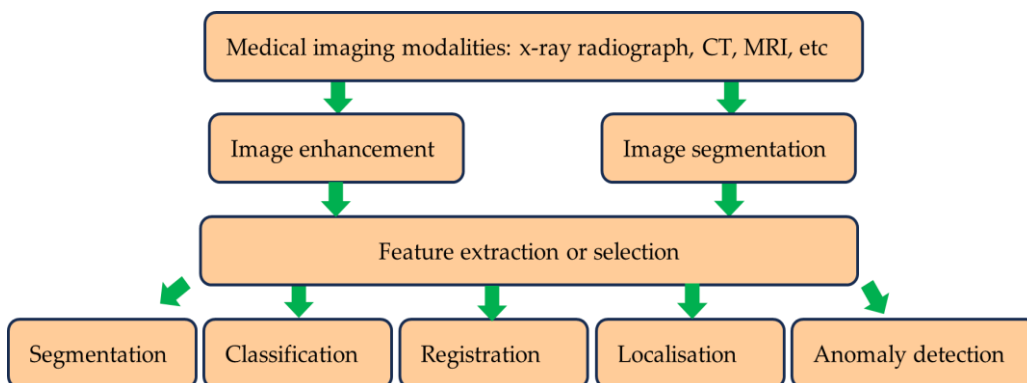


Figure 3. 1. A summary of deep learning usage with medical images (Shobayo & Saatchi, 2025)

3.1.3 Challenges in utilising deep learning in the medical field

Despite the significant advancements and capabilities of deep learning techniques in medical image analysis, there are certain limitations and challenges in their implementation and acceptance. For instance, most deep learning algorithms often lack explainability, i.e., they typically operate as black boxes (Takahashi et al., 2024). In the medical field where decision making processes are required for diagnosis and provision of treatment, the deep learning method of convolutional neural network (CNN), for instance, does not provide an insight into the manner

it came up with a decision. Other challenges that applications of deep learning techniques can face include:

- **Overfitting during training:** Overfitting causes poor accuracy when training deep artificial neural networks in recognising images not included during their training data set (i.e., the unseen or test images) even though the images in the training set could have been correctly recognised (Shorten & Khoshgoftaar, 2019; Srivastava et al., 2014; Tian, Y. & Zhang, 2022; Xiao, M. et al., 2021; Ying, 2019). It has several possible causes that include insufficient training examples to facilitate adequate generalisation and excessive model parameters for the architecture. However, there are techniques that could be valuable to deal with overfitting. A technique is known as drop out (Srivastava et al., 2014)(Garbin et al., 2020; Helmbold & Long, 2018; Piotrowski et al., 2020; Salehin & Kang, 2023) whereby some nodes in the architecture are temporarily left out during training. Another approach is to artificially extend the number of examples in the training data set through a process known as data augmentation (Kumar, T. et al., 2024; Xu, M. et al., 2023; Yang, Z. et al., 2023). A comparison of different image data augmentation methods was reported in (Nanni et al., 2021). The technique has been applied to mammograms (Oza et al., 2022) and CT images (Sandfort et al., 2019a). Augmentation can be performed by manipulating the images through processes such as kernel filters, geometric transformation, random erasing and mixing images. It can also be carried out through deep learning approaches such as adversarial training, neural style transformers and generative adversarial network (Shorten & Khoshgoftaar, 2019). An issue with performing data augmentation is selecting the best approach for a given set of images (Wei et al., 2023). An exploration of the influence of different data augmentation techniques on the explainability of deep learning methods was reported in (Liu, Xinyu et al., 2024).
- **Image annotation:** Many deep learning algorithms are supervised, i.e., they require labelled images indicating their categories during their training phase. The labelling requires annotation of images by qualified medical practitioners. Because deep learning requires large data for training, this process can be time consuming. However, there were reports of automated and interactive image annotation methods that can assist with this operation, e.g. (Galbusera & Cina, 2024a, 2024b; Jayaraj & Lokesh, 2023; Mamat et al., 2023; Zhang, Dengsheng et al., 2012; Zhang, Yu et al., 2024).

- Noisy images: Medical images can be noisy. Noise distorts the quality of images being used to train deep learning networks (Sriwong et al., 2021) and can reduce their ability to learn effectively.
- Interpretability: This is an area of great research interest to make decision making by deep learning artificial neural networks more transparent, i.e., moving away from a so called “black box” behaviour to more interpretable decision making. The issue of interpretability has been explored in many studies, e.g.,(Antamis et al., 2024; Carvalho et al., 2019; Jiang et al., 2024; Linardatos et al., 2020; Marcinkevičs & Vogt, 2023; Sriwong et al., 2021). Interpretability could be considered from multiple perspectives, e.g., user orientation for provided explanations, visualisation through graphs, charts rules etc, user comprehensibility through comprehensive reasoning, simplicity of explanations, local interpretation on a single data and global interpretation of overall data, consistency in explanations, transparency in decision making, and ethics and fairness in revealing bias and discrimination (Xu, B. & Yang, 2025).
- Data sharing complexities and small data sets: Medical data gathered by a single institution may be insufficient to allow effective training of deep learning algorithms and thus sharing across many institutions would be required. This however could be challenging due to the regulatory, technical and privacy concerns (Balachandar et al., 2020). Financial and time constraints as well as availability of required patients can limit the data set. However valuable techniques were devised to address small data set problems in deep learning (Piffer et al., 2024; Safonova et al., 2023).
- Ethical issues: There are numerous ethical issues requiring careful consideration, e.g. patient anonymity, consent/assent, data handling and protection, data sharing, and vulnerable participants (Chau, 2024; Chen, I. Y. et al., 2021; D’Antonoli, 2020; Safdar et al., 2020).
- Trust: There is an ongoing issue around relying on critical medical diagnostic results generated by when the manner of their generation is not sufficiently transparent (Sarmadi et al., 2024).
- Computational requirement and environmental issues: Training deep learning algorithms typically require high computational capabilities and long durations. Many general-purpose computers do not have the means of delivering the required computational resources and there is also the issue of the environmental aspects of using so much electrical energy to perform the required deep learning training (Chen, C. et al., 2020; Pietrołaj & Blok, 2024).

3.2 Convolutional neural network

CNN were the dominant technique for analysing medical images for diagnostic purposes (Masumoto et al., 2024). It has gained immense popularity since 2012 when high performance computing (HPC) became more accessible. This led to the ImageNet competition for different combination of the deep CNN networks to achieve better diagnostics results when compared to the human experts. CNN has been effective for several tasks such as image segmentation, detection, registration, localisation, and classification (Takahashi et al., 2024). They consist of numerous layers of convolutional filters with nonlinear activation functions, combined with pooling layers, dropout layers and fully connected layers. Their ability to extract complex spatial relationships and patterns in images has seen them used in various medical imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), X-ray radiographs, ultrasounds (US), histopathology and more recently in infrared thermal imaging (IRT) (Shobayo et al., 2024). The images associated with multiple diseases have been segmented, classified, registered and interpreted, ranging from bone fractures, cancer, liver diseases, pneumonia, Covid 19, etc. The CNN architecture applied to the discrete Fourier transformed infrared thermal images is shown in Figure 3.2. (Shobayo et al., 2024) Although the study was a pilot, the model was effective in screening wrist fracture in children.

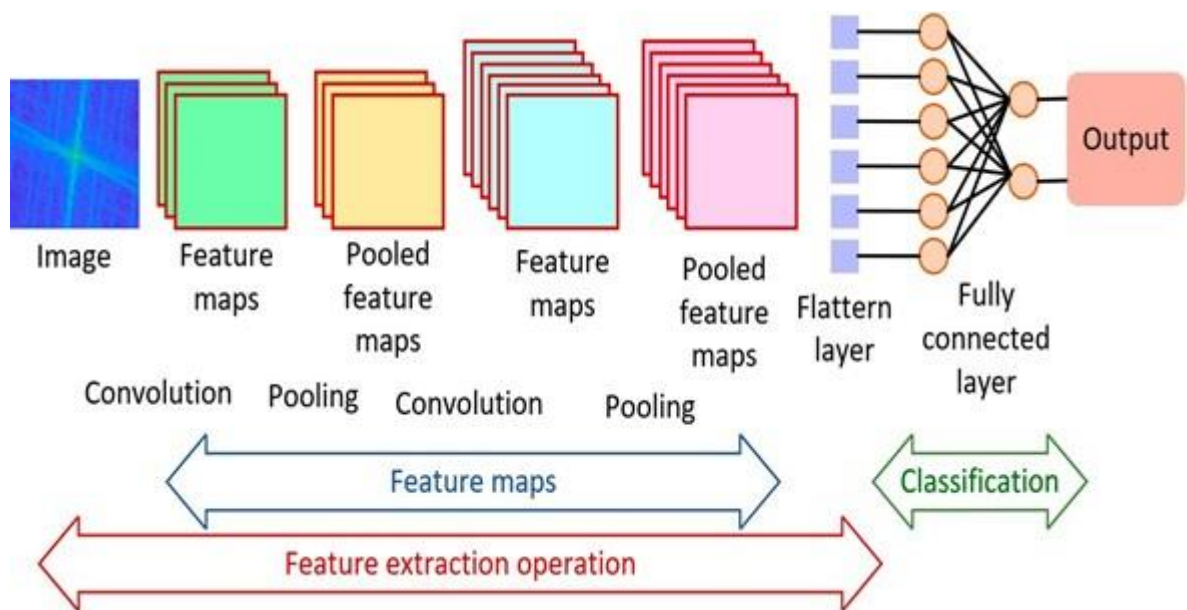


Figure 3. 2. An example of CNN used for the analysis of infrared thermal images (Shobayo et al., 2024).

The concept of convolution is simply a mathematical operation where a filter, also known as kernel, is applied on an input image for feature extraction. The convolution process is described

mathematically (Chua, 1997). Let the input image to the CNN be represented by a 3D matrix such that

$$X \in \mathbb{R}^{H \times W \times D} \quad 3.1$$

where H is the height of the image, W is the width of the image and D is the depth of the image (it represents the number of channels which is usually 3 for an RGB image). When applied to a set of convolutional filters for each input, the output of the convolutional filter can be represented by:

$$Z_{i,j,k} = \sum_{h=0}^{H_k} \sum_{w=0}^{W_k} \sum_{d=0}^D X_{i+h,j+w,d} \cdot K_{h,w,d,k} \quad 3.2$$

where $Z_{i,j,k}$ is the output feature map for the filter k at position (i, j) , K is the convolutional filter (kernel) H_k, W_k are the height and width of the kernel which is usually a sample of the size of the input. This process helps to identify local patterns such as the edges, shapes and textures from the input image.

When the CNN algorithm is used for classification purposes, it is combined with other layers such as the activation layers, pooling layer and the fully connected layer. The output from the convolutional filter is passed through an activation function $A_{i,j,k}$. If the rectified linear unit (ReLU) activation function is applied, the output from the convolution layer becomes:

$$A_{i,j,k} = \max(0, Z_{i,j,k}) \quad 3.3$$

This operation helps to add nonlinearity by ensuring only positive values are retained thereby helping to learn complex patterns from the image.

The pooling layer is then applied to the output of the activation layers. The effect of the pooling operation is to reduce spatial dimensions from the convolution operation thus help the network to capture small translations in the image. This process also helps to reduce the computational complexity of the network. Assuming the *max* (maximum) pooling function is applied to the activation layer, with a $p \times p$ window size, the output from the pooling layer is represented by:

$$A_{i,j,k} = \max_{h,w \in [1,p]} A_{i+h,j+w,k} \quad 3.4$$

The fully connected layer is connected to the output of a flattened *max* pooling layer output. This layer takes a vector value and is a standard neural network with each input connected to all the neurons in the next layer. The output of the fully connected (FC) layer is given by:

$$Z_l = W_l A_{l-1} + b_l \quad 3.5$$

where A_{l-1} is the input into the FC layer and W_l and b_l represent the weights and biases of the FC layer. The final or output layer which is also a vector is mostly passed through a SoftMax function for a classification task. This function helps to convert the output classes into probabilities with the following expression:

$$\hat{y}_i = \frac{\exp(Z_i)}{\sum_{j=1}^C \exp(Z_j)} \quad 3.6$$

Where \hat{y}_i represents the probability of predicting the class i , and C represents the total number of classes being differentiated. The full CNN network undergoes training to correctly learn the features of the image input. This training uses a loss function, usually a cross-entropy loss when considering a classification task. The cross-entropy loss is given by:

$$L = - \sum_{i=1}^c y_i \log(\hat{y}_i) \quad 3.7$$

where \hat{y}_i and y_i is the predicted probability for class i and the target label respectively.

The gradient of the loss function is calculated with respect to the weights through back propagation formula,

$$\frac{\partial L}{\partial W} = \frac{\partial L}{\partial Z} \cdot \frac{\partial Z}{\partial W} \quad 3.8$$

The weights are updated using the gradient descent, i.e.,

$$W_{new} = W_{old} - \eta \frac{\partial L}{\partial W} \quad 3.9$$

where η ($0 < \eta \leq 1$) represents the learning rate controlling convergence rate (Chua, 1997).

3.2.1 Literature review findings for CNN

Most studies have used CNN individually while there were studies that applied it in combination with another algorithm such as U-NET and GANs for medical image diagnostics. However, in this section the focus is on CNN application on its own or with some modifications (other CCN based algorithms, such as Transfer Learning is discussed in section 3.8). (Chittajallu et al., 2019) developed a simple CNN to classify fractures (fracture vs normal) on radiological images. Using a small dataset of ~200 images (hands, legs, ribs, neck), pre-processed by grayscale conversion and matrix representation, their model was trained in TensorFlow with a 90:10 train-test split

and binary cross-entropy. The architecture comprised convolution, max-pooling, flattening, and fully connected layers, with ReLU activations and mini-batch training. Reported average accuracies varied by region: hands 80.45%, legs 84.75%, ribs 80.65%, other 86.75%. Plots of epochs illustrated learning progression. The study indicates preliminary feasibility of CNN-based binary fracture detection, albeit on a small, heterogeneous dataset. Using CT scans of the Scapular of 90 patients, (Fang et al., 2025) built a Faster R-CNN model that improved fracture detection. Combining their model results with an orthopaedist prediction, they obtained an accuracy of 97.78%, beating orthopaedist-only (82.95%) and algorithm-only (92.05%) approaches; AUC was 0.866 versus 0.787 and 0.837. Specificity and positive predictive value increased most, with sensitivity and negative predictive value also higher, implying fewer misses and false alarms. Performance varied by site, with the scapular body, neck, and acromion most often identified. Limitations included small sample size and false positives from class imbalance. (Liao et al., 2022) proposed a human-guided attention regularization for CNN fracture classification, using scribble, bounding-box, or segmentation cues to steer CAMs. On scaphoid radiographs, study-wise accuracy rose from 76-80% (baseline) to 84-85% at optimal λ , with gains of 5-8%; performance matched ~84% human observers and avoided manual cropping. Annotation was fastest with scribbles (~24 s/study) versus boxes (31s) or segmentation (117s). On ankle radiographs, attention guidance improved study-wise accuracy by ~11% to ~86% on internal and external tests, outperforming self/residual/pyramid-attention baselines. The effectiveness of their model depends on tuning the regularization weight, and the method yields more faithful localisation alongside higher accuracy during training.

CNN was also used to classify other types of diseases besides bone fracture. For example, A lightweight CNN algorithm was used to detect Covid-19 from chest X-ray radiograph images (Chauhan et al., 2024). Their proposed CNN network was inspired by the ResNet model (Wu, Zifeng et al., 2019) in which all layers were not connected in sequence, creating a skip connection whereby neurons in a particular layer can be connected to another neuron further ahead. This arrangement helped to create a lightweight model that was effective in edge detection applications. Their study was compared with other CNN models such as CVDNet and deep GRU-CNN and showed a similar performance with the models compared with a reduced computational complexity. A deeper CNN in their work was proposed that used CT scans and X-ray radiographs for the detection of several pulmonary diseases including Covid-19 and viral pneumonia (Abdulahi et al., 2024). They used varied datasets, consisting of different image modalities for their work which provided another dimension of the efficacy of CNN in diagnostic imaging. They proposed a 26-layer deep CNN network which was inspired by a wide residual network

(WRN) (Zhang, Ke et al., 2017). It provided a faster training time when considering the deep nature of the architecture. This was achievable based on the sophistication of their hardware. Their model was effective in terms of the accuracy in detecting the different pulmonary diseases when compared to the traditional methods. A CNN model was used to classify ultrasonic images of fatty liver (Gao, Y. et al., 2022). A pertinent problem in their research was the similarity in the pathological ultrasonic images used for training the CNN algorithm. This sort of challenge might pose a problem for the CNN architecture as it may struggle to extract distinct features for the different pathological images. Therefore, there will be a need for deep convolution layers leading to computational complexities. They however used pixel-level feature extraction as a preprocessing step and then proposed a CNN architecture comprising of two convolutional layers, a pooling layer and a fully connected layer. They also experimented the proposed method with a skip connection and improved the accuracy when compared to other algorithms such as VGGNet.

3.3 Adaptive neuro-fuzzy inference systems

Adaptive Neuro-Fuzzy Inference Systems (ANFIS) merge the human-readable rule structure of fuzzy logic with the adaptive learning power of neural networks, yielding a versatile framework for modelling complex, nonlinear relationships. ANFIS is structured into five sequential processing layers which is often labelled the fuzzification layer, rule-strength (product) layer, normalisation layer, consequent (de-fuzzification) layer, and overall output layer with each layer corresponding to a key step in a Takagi-Sugeno-style fuzzy inference process (Jang, 1993). Figure 3.3 (Walia et al., 2015) shows the basic structure of an ANFIS architecture and each layer with the corresponding formula is provided below.

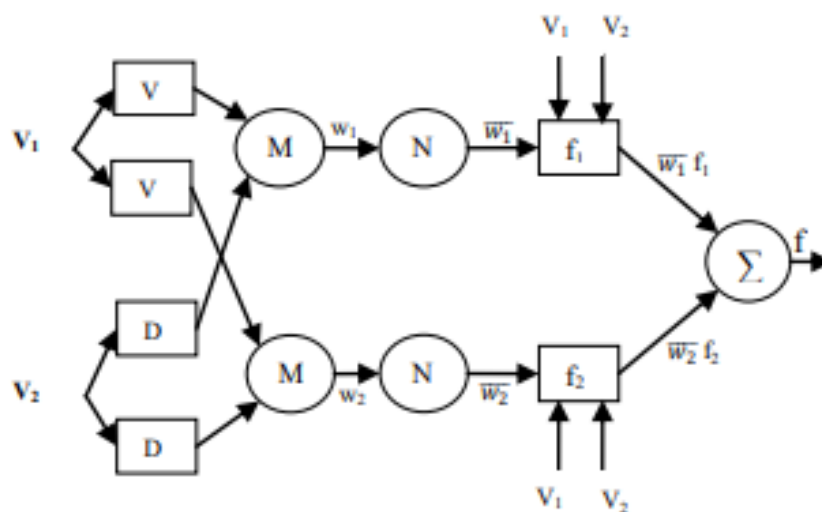


Figure 3. 3. ANFIS Basic architecture (Walia et al., 2015)

Layer 1: Fuzzification. Each crisp input (e.g., variables v and d) is passed through a set of parameterised membership functions (MFs), typically bell-shaped or Gaussian. These “premise” parameters, i.e., centres and widths are initially guessed and later tuned during training to capture the underlying data distribution.

Layer 2: Rule Strength. For each fuzzy rule, the membership degrees from Layer 1 are combined, usually via multiplication to produce a firing strength w_i . Take for example if we have an ANFIS system with two inputs and two rules:

Rule 1: “If v is V_1 and d is D_1 , then $f_1 = p_1v + q_1d + r_1$ ”

Rule 2: “If v is V_2 and d is D_2 , then $f_2 = p_2v + q_2d + r_2$ ”

Here, V_1, D_1, V_2, D_2 , are the nonlinear MF parameters (antecedents), and p_i, q_i, r_i are linear “consequent” coefficients.

Layer 3: Normalisation. Each raw firing strength w_i is normalised by the sum of all w_j , yielding $\bar{w}_i = \frac{w_i}{\sum_j w_j}$. This step ensures the rule outputs are weighted proportionally.

Layer 4: Consequent Computation. The normalized strengths \bar{w}_i multiply their respective linear consequents f_i . Because each f_i is already a linear function of the original inputs, this layer effectively performs de-fuzzification.

Layer 5: Output Aggregation. The system’s final output is the sum of all weighted consequents:

$$y = \sum_i \bar{w}_i f_i(x) \quad 3.10$$

To arrive at an optimal ANFIS, both the premise (MF shapes) and consequent parameters must be identified from data. ANFIS employs a hybrid learning algorithm combining least-squares estimation (for the linear consequents) with gradient-descent backpropagation (for the nonlinear MF parameters) (Jang, 1993). In each iteration, the forward pass fixes the MFs and solves for the best-fit linear consequents via least squares; the backward pass then propagates the output error to adjust the MF centres and widths by steepest descent. This two-phase procedure accelerates convergence by reducing the dimensionality of the search space, compared with pure gradient methods alone.

By integrating these techniques, ANFIS automatically crafts an interpretable fuzzy rule set that emulates the training data and generalises to unseen cases with minimal error. Its capacity to both

learns from examples and express knowledge as “if-then” statements has made ANFIS widely popular in fields such as control, system identification, forecasting, and classification (Kar et al., 2014; Walia et al., 2015). This synergy of neural learning and fuzzy reasoning explains why ANFIS remains a cornerstone technique in the development of intelligent, data-driven decision-support systems (Karaboga & Kaya, 2018).

3.3.1 Literature review findings for ANFIS

Automated diagnostic platforms are increasingly utilised across diverse types of medical data, ranging from physiological signals to imaging (Azeez & Abdulazeez, 2025; Pilehvari et al., 2024; Purwono et al., 2025; Shobayo & Saatchi, 2025). ANFIS has been successfully employed in numerous medical related applications where medical signals and images serve as the primary inputs to the decision models. Typical examples include detection and diagnosis of diabetes, blood pH imbalances, valvular and rheumatic heart conditions, epileptic seizures, prostate malignancies, and various cancers (colon, leukaemia, lymphoma) via microarray data. ANFIS has also been applied to ophthalmic and optic nerve disorders, analysis of Doppler ultrasound signals (including internal carotid assessments), interpretation of electroencephalogram (EEG) recordings, and identification of arterial abnormalities in the eye. These studies collectively demonstrated the versatility and effectiveness of ANFIS in medical decision-support systems (Avci & Turkoglu, 2009; Hosseini & Zekri, 2012; Übeyli, 2009). ANFIS models were widely used in literature for classification of medical images mainly due to its fast convergence and effective classification abilities even with smaller datasets. For example, the study by Hemanth and colleagues (Hemanth et al., 2011) applied an ANFIS to classify four types of abnormal magnetic resonance imaging brain-tumour images which included metastases, meningioma, glioma, and astrocytoma, using six texture features derived from grey-level co-occurrence matrices. Each input feature was fuzzified with two generalised bell-shaped membership functions, yielding 64 fuzzy IF-THEN rules. The model’s training combined least-squares estimation (for rule consequents) with gradient-descent tuning (for membership parameters) over 200 iterations. On a 460-image dataset (120 training, 340 testing), ANFIS achieved an overall classification accuracy of 93.3% which was higher than a fuzzy-nearest-centre classifier (88.6%) and a back-propagation neural network (85.7%). Moreover, the ANFIS model converged in just 1,540 CPU cycles, roughly one-tenth the time required by the other methods, while producing low mean-square errors (training ≈ 0.001 , testing ≈ 0.15). Kumar and colleagues (Kumar, R. et al., 2023) presented an ANFIS-based approach to predict COVID-19 epidemic peaks and infection counts in India. They combined nationwide case data sourced from cloud repositories

with local demographic and health indicators which includes population density, age distribution, comorbidities, and infrastructure metrics, to construct a two-input, single-output Sugeno model. Using nine trapezoidal membership functions per input and 81 fuzzy rules, the model was trained via a hybrid least-squares and gradient-descent algorithm. Validation against unseen data yielded a low mean square error ($MSE = 1.184 \times 10^{-3}$) and an overall predictive accuracy of 86%, outperforming linear and multiple-regression baselines ($\approx 83\%$) both in accuracy and computational efficiency (438 s versus 540–720 s). A hybrid ANFIS framework optimised by Adam and Particle Swarm Optimisation (PSO) was proposed to improve Parkinson’s disease (PD) diagnosis accuracy (Pasha et al., 2024). Using a public UCI dataset of 756 voice recordings (755 features), they first employed an Extra Trees ensemble to select the top five predictive features. Two ANFIS models were then trained separately: one with Adam (gradient-based) and one with PSO (swarm-based) optimizers. The PSO-tuned ANFIS achieved lower training loss and higher precision, while the Adam-tuned model yielded superior accuracy, F1-score, and recall. Across varying epochs, membership functions, and PSO particle counts, both models converged efficiently, with PSO requiring fewer iterations but Adam delivering slightly faster convergence. The best configuration (1000 epochs, 50 particles, four rules per feature) produced test accuracies above 84%, precision up to 91%, and F1-scores near 84%.

These findings demonstrate that ANFIS can effectively leverages both neural learning and fuzzy interpretability to deliver rapid, accurate tumour classification in MRI imaging (Hemanth et al., 2011), has the capacity to deliver timely, interpretable guidance for public-health policy and sustainable strategic resource allocation and planning (Avci & Turkoglu, 2009; Hemanth et al., 2011; Hosseini & Zekri, 2012; Übeyli, 2009), and when combined with adaptive optimizers, it can produce a better interpretable strategy for early PD detection in clinical settings (Pasha et al., 2024). Other studies that have used ANFIS for the classification of medical images is presented in the Table 3.1.

Table 3. 1. Summary of ANFIS technique for medical image analysis

Article	Imaging Modality	Disease/Body Part	ANFIS Variant Used
(Al-Ali et al., 2021)	X-Ray	COVID 19	ANFIS
(Sharma & Mukharjee, 2012a)	MRI	Brain Tumour	GA-ANFIS
(Birgani et al., 2019)	MRI	Brain Tumour	Enhanced ANFIS
(Chatterjee & Das, 2020)	MRI	Brain Tumour	Enhanced ANFIS

(Kumarganesh & Suganthi, 2023)	MRI	Brain Tumour	ANFIS
(Richard et al., 2025)	RGB	Down Syndrome	ANFIS-CNN
(Balakrishnan & Perumal, 2025)	OCT	Kidney microanatomy	ANFIS-CNN
(Mayeta-Revilla et al., 2025)	MRI	Brain Tumour	ANFIS
(Tiwari et al., 2025)	MRI	Brain Tumour	Deep Belief-ANFIS

3.4 Recurrent neural network

Recurrent neural networks (RNNs) are another class of neural networks that have gained significant research consideration for modelling of sequential data. Their variability of the length of their input and output also makes them suitable for natural language processing tasks (Jeong et al., 2022). A RNN has an internal memory unlike feedforward networks which helps to keep memory of the hidden states across different time stamps. In RNN, there is a feedback loop between the outputs of the hidden layers (Kim, J. et al., 2018). This arrangement allows the RNNs to learn sequential patterns, making them suitable for tasks such as time series prediction and video analysis (Zhang, Huanhuan & Qie, 2023). RNN architecture normally suffers from the issue of vanishing gradients (Rajeev et al., 2019). Their use transcends sequential and textual data. They can also be applied to image modalities with time-series characteristic information such dynamic imaging in functional MRIs to monitor progress of disease in a patient and to check how they respond to treatments. This functionality has been made popular by a variant of the network known as Long Short-Term Memory Network (LSTM). The LSTM works by introducing self-loops to allow for the flow of gradients for long durations (Yao et al., 2024). The recurrent structure of the neural network is always enforced by the LSTM by introducing gating functions on the neurons in the hidden layer. The architecture of the RNN is shown in Figure 3.4 (Zhang, Huanhuan & Qie, 2023).

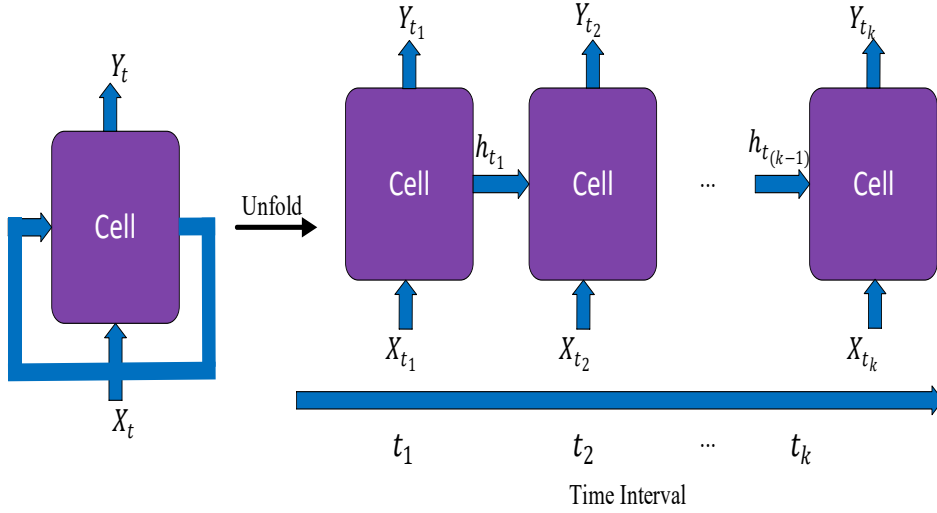


Figure 3. 4. Structure of the RNN (Zhang, Huanhuan & Qie, 2023).

The mathematical representation of RNN is described next (Kani & Elsheikh, 2017). Let the sequential input to the RNN at time step t be x_t where $x_t \in \mathbb{R}^n$ is a vector. Hence the input sequence becomes

$$x_1, x_2, \dots, x_T \quad 3.11$$

where T represents numbers of time steps in the sequence. For every time step t , a hidden state (h_t) is maintained. Each hidden state is updated as

$$h_t = f(W_{hh}h_{t-1} + W_{xh}x_t + b_h) \quad 3.12$$

where W_{hh} represents the hidden states or recurrent weights, W_{xh} is the weight of the connection between the input layer and the hidden layer and b_h is the bias for the hidden state. The function f represents the activation function, typically the ReLU or the hyperbolic tangent. To obtain the output of the RNN at each time step, the output from each hidden state is multiplied by its associated weights and added with a chosen bias vector. This is represented as

$$y_t = g(W_{hy}h_t + b_y) \quad 3.13$$

where W_{hy} represents the weight matrix between the hidden state and the output, and b_y represents the output bias vector. The function g is the activation function for the output, which could be task dependent. SoftMax activation function is mostly employed for classification tasks (Kani & Elsheikh, 2017).

3.4.1 Literature review findings for RNN

RNN is widely used for machine learning or deep learning tasks that are time related. For image diagnosing which also include bone fracture, it can be used to monitor the disease progression and patient response to treatments, based on different time stamp that the images are being taken. For example, 3D wrist ultrasound images from 30 children (55 scans; 19 fractures) were used by (Hareendranathan et al., 2023) to develop a bone probability map, which included local phase, feature symmetry, integrated backscatter using recurrent models (RNN/LSTM/GRU). LSTM achieved the highest accuracy (90%), while GRU provided the best sensitivity (86%); all RNNs exceeded 80% accuracy and used far fewer parameters than AlexNet, VGG, ResNet, and Inception baselines. Incorporating the probability map improved accuracy, specificity, PPV and NPV across models. Per-patient evaluation detected 6/7 fractures with two false positives. The method runs near-real time (<1 s/volume) and could support paediatric ED triage.

(Jothi & Jayanthi, 2024) introduced a deep multi-filtering pipeline: histogram equalisation and guided filtering for enhancement; a dilated Mask R-CNN augmented with recurrent Kalman and fuzzy filters for segmentation; and a self-attention RNN that classifies concatenated backbone and LBP/HOG/SURF features. On a 10,000 x-ray radiographs with 70/30 split, their method surpassed CNN, Faster R-CNN, LSTM, and SVM baselines, by achieving accuracy 0.978, precision 0.981, sensitivity 0.975, specificity 0.980, F1 0.978, MCC 0.955, with lower FPR/FNR (0.020/0.025). Segmentation also improved, indicating better edge/region fidelity and defect visibility than comparators. Findings support accurate, scalable fracture triage.

Other works that have based their classification task on RNN, typically combined it with CNN for the feature extraction stage for example, 192 wrist radiographs consisting of 111 fracture and 82 normal, was used by (Rashid et al., 2023) to propose a DCNN-LSTM pipeline with preprocessing, adaptive histogram equalisation, and rotation-based augmentation to address class imbalance. On non-augmented data, DCNN achieved 84.48% accuracy: adding LSTM improved accuracy to 86.21%. With augmentation, DCNN reached 86.54% and DCNN-LSTM achieved 88.24% accuracy, with sensitivity 92.17%, specificity 82.93%, F1-score 90.0%, and Cohen's kappa, κ 75.7%. These results indicate that fusing dilated CNN features with bi-LSTM classification and balancing classes improves wrist-fracture detection over DCNN alone, supporting computer-aided triage. They noted smaller datasets as a limitation and advocate larger, multi-institutional datasets to further optimise architecture and generalisation performance.

A summary of works that have used RNN for other diseases classification are shown in Table 3.2.

Table 3. 2 Recurrent neural network applications.

Author	Image Modality	Task	CNN feature extraction	Disease/body part	Variant used
(Cui & Liu, 2019)	MRI	Classification	Y	Alzheimer's	BGRU
(Anbalagan & Balasubramanian, 2024)	Histopathology	Classification	N	breast cancer	None
(Amarneni & Valarmathi, 2024)	MRI	Classification/Segmentation	N	Brain tumour	LSTM
(Yao et al., 2024)	MRI	Segmentation	Y	Aorta	LSTM
(Zhu, Kaiyue et al., 2022)	MRI	Classification/Localisation	Y	Knee ligament	LSTM
(Gulshan & Arora, 2024)	IRT	Classification	Y	Diabetes mellitus	LSTM
(Rajeev et al., 2019)	CT	Image Denoising	N	Lungs	LSTM
(Ayub et al., 2022)	MRI	Registration	Y	Brain Cancer	LSTM

3.5 Autoencoders

Autoencoders are unsupervised learning models used for dimensionality reduction and feature extraction with minimal distortion when their input is compared to their output (Zhang, Yifei, 2018). They play an important role in the deep learning paradigm for medical image analysis (Baldi, 2012). They can help denoise or compress medical images and are useful in anomaly detection, where unusual patterns in images indicate potential medical issues. They can also be used as a semi-supervised deep learning model to produce annotated data in situations where there is a lack of substantial amount of annotated dataset available for training any deep learning network for tasks such as classification or segmentation (Vorontsov et al., 2022). Their architecture consists of an encoder and decoder structure with a latent space to store the value of the compressed data. Both the encoder and the decoder comprise of a fully connected feedforward neural network. The encoder converts the input image into a low dimension compressed version, which is referred to as latent space or the encoder. The latent space contains only essential features of the input from the encoder and is kept as shallow as possible in terms

of the number of neurons used to retain the compressed version of the input and computational efficiency. The encoder in turn transforms the latent space to a reconstruction of the input. A loss function is generally used during training to compare the input with its reconstruction (Wang, Wei et al., 2014). The architecture of an autoencoder with an MRI image as its input is shown in Figure 3.4 (Shobayo & Saatchi, 2025).

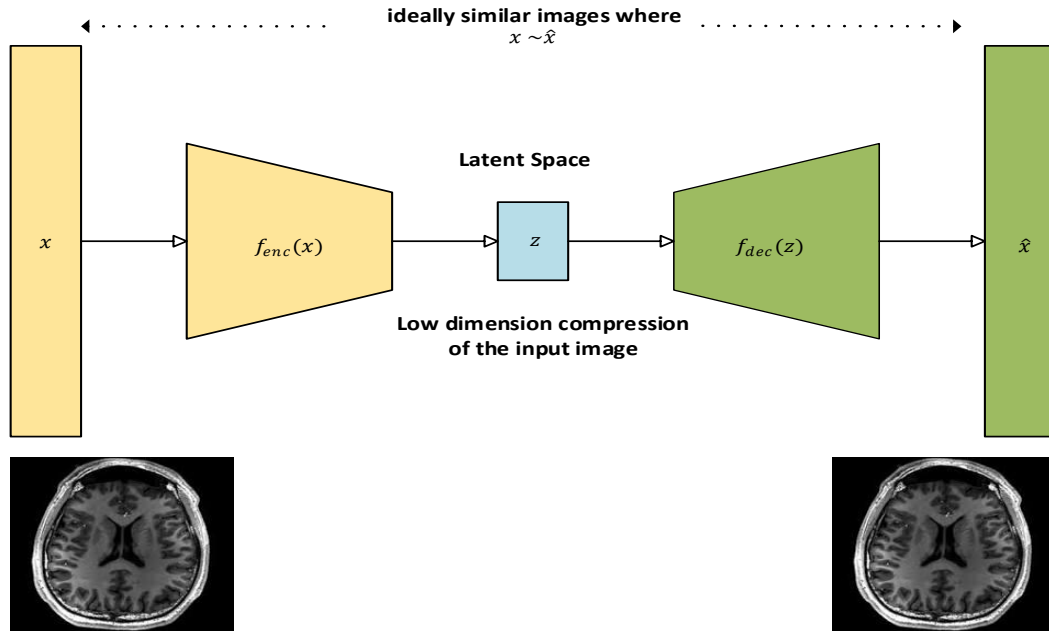


Figure 3.4. An Autoencoder architecture reproducing an MRI image (Shobayo & Saatchi, 2025).

The mathematical representation for the dimensionality reduction function of the autoencoder's encoder and decoder is described next (Skansi, 2018). Let the input image to the encoder be represented by

$$x \in \mathbb{R}^n \quad 3.14$$

The encoder in turn transforms the input x into a lower-dimensional latent representation

$$z \in \mathbb{R}^d \quad 3.15$$

where $d < n$.

The transformation function that produced the latent information is given as:

$$z = f_{enc}(x) = \sigma(W_{enc}x + b_{enc}) \quad 3.16$$

where $W_{enc} \in \mathbb{R}^{d \times n}$ is the encoder weight matrix, $b_{enc} \in \mathbb{R}^{d \times n}$ represented the encoder's bias vector and σ represents the selected activation function of the encoder which is commonly ReLU

or sigmoid function. The latent representation z of the input x provides compressed information by only capturing the essential features of the input image, a feature referred to as bottleneck in which the dimension of the latent feature is smaller than the input vector. This facilitates the autoencoder to learn a better way to efficiently represent the input vectors. The decoder network tries to transform the latent space z back to the reconstructed input \hat{x} to match the input vectors x . The decoder function can be represented as thus:

$$\hat{x} = f_{dec}(z) = \sigma'(W_{dec}z + b_{dec}) \quad 3.17$$

The loss function is used to minimise the difference between the input x and the reconstructed input \hat{x} . This difference is normally quantified by binary cross-entropy error (BCE) when the dataset is binary or Mean Squared Error (MSE) for multi-class data.

The MSE loss computed with the following function

$$L(x, \hat{x}) = \frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2 \quad 3.18$$

And the BCE loss is computed as follows

$$L(x, \hat{x}) = - \sum_{i=1}^n [x_i \log(\hat{x}_i) + (1 - x_i) \log(1 - \hat{x}_i)] \quad 3.19$$

Training of the autoencoder requires the finding of the parameters for the weights and biases of the encoder and decoder such that the reconstruction loss is minimised over the training data. Gradient based optimisation techniques such as stochastic gradient descent (SGD) or Adam can be used for this purpose (Skansi, 2018). Summary of works that have used autoencoder is shown in Table 3.3.

Table 3.3. Autoencoder based techniques.

Author	Image modality	Task	Disease/ body part
(Vorontsov et al., 2022)	MRI	Augmentation/segmentation	Brain
(Juneja et al., 2021)	MRI	Denoising	Prostate
(O' Sullivan et al., 2022)	CT + others	Classification	Face
(Wolf et al., 2023)	CT	Augmentation	Various

(Chen, R. et al., 2021)	MRI & CT	Classification	Intracerebral haemorrhage
(Shvetsova et al., 2021)	X-ray/ Histopathology	Anomaly detection	Various
(Elhassan et al., 2023)	Single-cell images	Classification	Myeloid Leukaemia
(Zhang, Haibo et al., 2022)	None	Anomaly detection	None
(Li, D. et al., 2020)	CT	Classification	Covid-19
(Zhang, Huilian et al., 2024)	MRI	Denosing/classification	Autism/Brain

3.5.1 Literature review findings for Autoencoder

Autoencoders are associated with both unsupervised and semi-supervised deep learning tasks. These tasks are normally preferred in the absence of ample annotated datasets required for deep learning activities especially in imaging analytics. Autoencoders are used to augment small number of annotated datasets often before a segmentation task. For example, a method known as GenSeg was proposed which combines the generative aspects of the autoencoders to generate the latent representation of the tumour cells from a labelled health image and using U-Net architecture to obtain the unique information of tumours present in the MRI images. There were also other related studies (Wolf et al., 2023). Noise reduction is another task that has benefitted from generative features of the autoencoder framework. The fusion of Bayes shrinkage fused wavelet transform (BSbFWT) was proposed for noise removal and an auto encoder block for generating a noiseless variant of an MRI image of prostate cancer (O' Sullivan et al., 2022). As MRI images can be prone to Gaussian and Rician noise which is introduced during image capturing by the MRI device and the imaging environment. The effectiveness of their noise reduced generated images is measured using several parameters like values of peak signal to noise ratio (PSNR), mean squared error (MSE), structural similarity index metric (SSIM) and mean absolute error (MAE) which outperforms numerous traditional filtering approaches presented in their work. Auto encoders can also be used in classification tasks (Chen, R. et al., 2021; Elhassan et al., 2023; Li, D. et al., 2020; O' Sullivan et al., 2022).

Using X-ray images from 87 subjects (39 osteoporosis, 38 controls), (Nasser et al., 2017) proposed a three-stage pipeline for osteoporosis prediction which includes adaptive histogram equalisation, patch-based feature learning with a stacked sparse autoencoder (SSAE) and pooling which is followed by an SVM classification. Ten-fold cross-validation showed the linear-kernel

SVM outperformed polynomial, RBF, and sigmoid, achieving 95.50% accuracy versus 75.56-82.48% for alternatives. The approach addresses small datasets by learning high-level representations and supports semi-supervised training with unlabelled data. Results indicate SSAE features plus SVM can discriminate osteoporotic from control cases on radiographs, though larger multicentre validation is needed. (Nguyen et al., 2023) presents an automated pipeline to estimate hip bone mineral density (BMD) from radiographs by suppressing soft-tissue effects with a convolutional autoencoder trained on synthesized soft-tissue images, detecting landmarks with a decentralised CNN, and regressing grayscale features from three hip regions. On 939 cases from two hospitals, autoencoder-enhanced images improved agreement with DXA (fit $r=0.83$; 10-fold $r=0.81$, MAE 0.069 g/cm²) versus original radiographs ($r=0.72$, MAE 0.083) and raised success rates for errors ≤ 0.045 g/cm². Processing time was ~ 1.4 s/case. Findings suggest a low-cost alternative for BMD screening where DXA is scarce, though fixed image-acquisition settings and dataset scope may limit generalisation. (Kim, D. et al., 2023) developed an unsupervised stacked autoencoder (SAE) for limb fracture prediction using bilateral symmetry and low-frequency transverse vibration. Differences between frequency response functions of a “reference” healthy limb and the test limb are converted to STFT “virtual spectrograms” for classification. On plastic pipes, baseline accuracy was 90.5% with crack–healthy confusion; fine-tuning raised accuracy to 100%, and external validation improved from 80% to 93.3%. Cadaver tibia tests showed lower eigenfrequencies and higher responses in fractured states, enabling clear separation. In a patient case, the healed leg (with intramedullary nail) displayed responses similar to no-fracture, and the SAE correctly distinguished fracture, healing, and healthy conditions. In a second patient with bilateral fractures, four classes (fracture; healing 1-month; healing 6-months; no-fracture) were classified with 100% accuracy. The approach achieved $>93\%$ accuracy overall but does not yet quantify fracture size and relies on reference signals.

Autoencoder were also used to detect complex anomalies presented in medical imaging (Shvetsova et al., 2021; Zhang, Haibo et al., 2022; Zhang, Huilian et al., 2024). Autoencoders were applied to chest X-ray and digital pathology images (Elhassan et al., 2023). Abnormalities that are barely visible such as metastases in lymph nodes always proved difficult to detect as they resemble normal images in pathological slides. They proposed a deep perpetual autoencoder that learnt the shared patterns of normal images and content similarities with abnormal ones and restores them correctly. For evaluation of their mode, they used the receiver operating characteristics (ROC) as it integrates the classification performance of the normal and the abnormal class. Their models were also evaluated on non-medical related images. Their model

performed well with medical images when compared to non-medical images making autoencoder suitable in medical imaging climes.

3.6 Generative adversarial network

Generative adversarial networks (GANs) are groups of deep learning artificial neural networks that can be used for generating synthetic medical images, data augmentation, and improving image resolution. They are valuable for enhancing small datasets, which are common in medical imaging, and can also be used to create better training data for improving model performance. GAN uses an unsupervised learning algorithm and can be used for mostly semi-supervised and unsupervised learning (Gao, J. et al., 2022; Singh & Raza, 2021). A GAN network consists of two main parts, namely the generator (G) and the discriminator (D). The generator, which comprises of a MLP, learns the data distribution of the input image and produces a similar image to the input, also known as “fake data”. The job of the discriminator, which is also an MLP, is to discriminate between the generated image from the generator network and the input image. The result of this discrimination, which constitutes an error between the original input image and the generated image is fed back to the generator input to make the generated image more realistic i.e., closer to the original image. During training, the weights of the generator and discriminator are alternately updated, and the weights updates of the generator come from the discrimination error. Both networks are engaged in a competing optimisation process. This process continues until there is an equilibrium between the generator and the discriminator networks (Jeong et al., 2022; Xun et al., 2021; Yi, X. et al., 2019). The architecture of the GAN network is shown in Figure 3.5 (Singh & Raza, 2021).

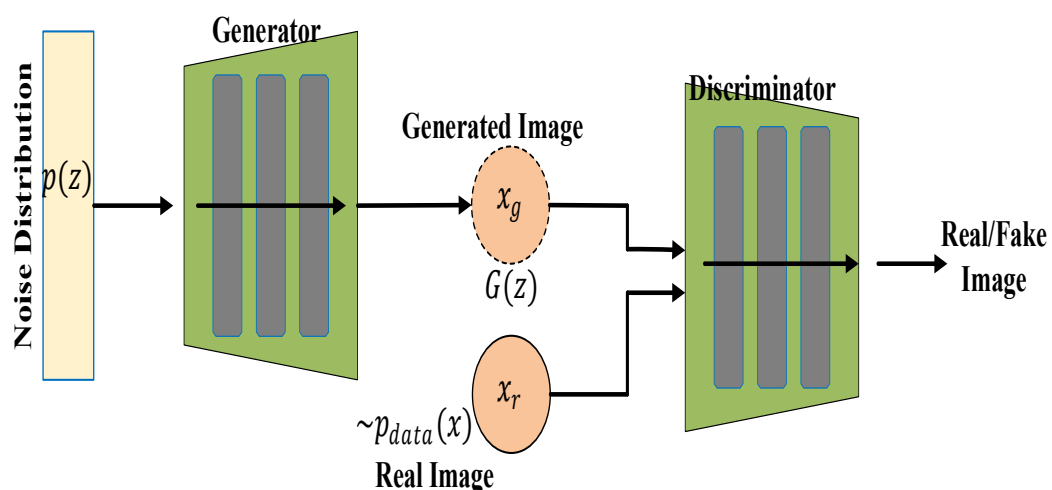


Figure 3. 5. Structure of the GAN network (Singh & Raza, 2021).

The Generator (G) consists of an input noise vector z such that $z \sim p_z(z)$, where p_z is a prior distribution which could be Gaussian (Wang, Yang, 2020). The output of the generator, which could be an image is represented by $G(z)$ and the objective is to maximise $\log(D(G(z)))$. The input to the discriminator (D) is the image data sample x and the output of the generator network. The output to the discriminator is the probability that x is real or fake. This is represented by $D(x) \in [0,1]$, where 0 represents fake and 1 represents real. The goal of the discriminator network is to maximise $\log(D(x) + \log(1 - D(G(z))))$. The discriminator loss \mathcal{L}_D is given by

$$\mathcal{L}_D = -\mathbb{E}_{x \sim p_{data}}[\log D(x)] - \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))] \quad 3.20$$

where $-\mathbb{E}_{x \sim p_{data}}[\log D(x)]$ maximises the probability of correctly classifying the real data and $\mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))]$ minimises the probability of misclassifying the generated data (Wang, Yang, 2020). When both generator and discriminator are at a stable equilibrium, the combined minmax game function is defined as

$$\text{Min}_G \max_D \mathbb{E}_{x \sim p_{data}}[\log D(x)] + \mathbb{E}_{z \sim p_z}[\log(1 - D(G(z)))] \quad 3.21$$

Summary of works that have used GAN for medical image analysis is provided in Table 3.4

Table 3.4. GAN based techniques.

Article	Image Modality	Task	Disease/body part	Variant used
(Gao, J. et al., 2022)	MRI/Retina fundus	Image synthesis	-	-
(Hamghalam et al., 2020)	MRI	Image resolution	Brain	Cycle-GAN
(Zhu, Qi et al., 2021)	CT	Image synthesis	Covid	Enhanced vanilla
(Li, Y. et al., 2021)	X-Ray/CT	Image Denoising	Chest/Thorax	CGAN
(Ahmad et al., 2022)	Various	Image resolution	Various	Enhanced vanilla
(Mutepfe et al., 2021)	-	Image synthesis	Skin cancer	DCGAN
(Touati et al., 2021)	MRI/CT	Image synthesis	Head/Neck	Vanilla GAN
(Xiao, Y. et al., 2023)	MRI/CT	Image resolution	Bladder cancer	Enhanced Vanilla
(Mahapatra et al., 2019)	Retina Fundus/MRI	Image resolution	Various	Vanilla GAN
(Uzunova et al., 2020)	CT/MRI	Translation	Thorax/brain	CGAN

3.6.1 Literature review findings for GAN

Deep learning technique such as GANs can be used in medical imaging in two different ways. The generative network in GANs has been used for image synthesis purposes to generate synthetic datasets for tasks where there are limited annotated datasets and using the discriminating networks of GANs for anomaly detection (Singh & Raza, 2021). The quality of the generated synthetic images can be measured by the use of some qualitative metrics such as Fréchet inception distance (FID) which is a measure of similarities between the representations of the generated and the real input images, structural similarity index measure (SSIM), which indicates the similarities of the structures (usually image contrast and brightness) and peak signal to noise ratio (PSNR), which are used to analyse the sensitivity of the generated image. Sometimes the generated images were also given to the surgeon for comparison with the original image and measured with the Fleiss kappa coefficient (Sindhura et al., 2024; Zhao et al., 2025). PSNR is the most important metrics when dealing with medical images. Quantitative metrics used for generated image quality includes number of parameters (NoP), to represent the total number of trainable parameters in the GAN network, and floating points of operations (FLOPs), which measures the cost of computation of the network (Gao, J. et al., 2022). Most studies that have used GANs have used it for image synthesis, image resolution, image translation, and image denoising. For example, (Sindhura et al., 2024) used CT images of sub-axial spine fractures to train an extended DCGAN to synthesize fracture images and clinically validate them via two Visual Turing Tests with three spine surgeons. The surgeons struggled to distinguish real from synthetic images (mean accuracy $\approx 50\%$; Cohen's Kappa $k = 0.336$) and showed substantial agreement when identifying fracture types on generated images ($k=0.612-0.769$). Augmenting training with DCGAN images improved type identification: a VGG16+ResNet50 ensemble rose from 82.14% to 84.23% accuracy, outperforming single backbones ($\leq 72.14\%$ without augmentation). Results support GAN-based augmentation and ensemble learning to mitigate imbalance and enhance clinical spine fracture triage. By retrospectively analysing 1,063 anteroposterior (AP) hip radiographs, (Mutasa et al., 2020) trained a two-stage CNN (localisation then classification) to detect and grade femoral neck fractures (normal, Garden I/II, Garden III/IV). With advanced augmentation based on digitally reconstructed radiographs (DRR) from CT and GAN-generated images, the model achieved 92.3% accuracy (AUC 0.92) for fracture vs. no fracture, and 86.0% accuracy (multi-class AUC 0.96). DRR augmentation alone raised AUC

to 0.91, GAN alone to 0.87, versus 0.80 without. Class-wise, sensitivity/specificity were 0.54/0.93 for Garden I/II and 0.91/0.83 for Garden III/IV; misses were typically off by one class.

For denoising tasks, GANs were used for denoising X-ray images using the CGAN variant of the GAN architectures (Li et al., 2021). They purposed to deal with the spatially varying noise, which is often overlooked when dealing with medical images. To achieve this, the gradient of original image was merged with the noisy image to obtain the conditional information for the CGAN network thereby enhancing the contrast. The convolutional layers of the generator were used in full for better feature extraction. For improved consistency between the real and fake images, the reconstruction loss was combined with Wasserstein GAN (WGAN) loss to create an objective loss for the network. They obtained remarkable PSNR and SSIM performance when compared to other state-of-the-art GAN architectures.

3.7 U-Net

U-Net architecture combines the best of CNN and encoder-decoder models, specifically for the purposes of segmenting medical images (Azad et al., 2024; Ronneberger et al., 2015). They have found applications in major medical image modalities such as CT, X-ray radiograph and MRI. The U-Net's ability to exploit small, annotated data samples (based on its fully connected layers) by leveraging data augmentation and improved feature extraction made it a valuable technique for medical image segmentation (Azad et al., 2024). The U-shaped architectures with skip connections help to delineate objects in images, making it highly effective in medical image analysis, particularly in tasks like tumours detection, organ delineation, and segmentation of medical images from various modalities and they have been widely embraced variants among the many different deep learning networks (Krithika Alias Anbudevi & Suganthi, 2022).

U-Net architecture is mainly composed of two paths. The first path is referred to as the contracting or encoder path. It uses a down sampling module that consists of several repeating convolutional blocks for semantic and contextual feature extraction. Each convolution block has two successive 3x3 convolutions, ReLU activation functions and the pooling layer (Azad et al., 2024). The pooling layer serves to increase the receptive field of the convolution network with no extra burden of computing resources that might be introduced by an additional convolution block. The second path of the U-net is the expansive path, or the decoder path and it is saddled with the task of up sampling spatial resolutions of the feature maps from the contracting path, usually by a factor of two. During this operation, the dimensions of the features are reduced, and a pixel-wise classification/resolution score is produced. The expansive path is made up of a 2x2 transposed convolution layer (reversing the operation in the contracting path), which is followed

by a 3x3 convolutional layer and a ReLU activation function. There is also a bottleneck layer which serves as a connection between the two paths. It is also comprised of two blocks of 3x3 convolution layers and a ReLU activation function. The embedded skip connections in the bottleneck copy the output of each stage of the paths, helping to learn contextual and semantic representations in the deep and shallow layers respectively (Azad et al., 2024; Ehab & Li, 2023). The U-Net architecture is shown in Figure 3.6 (Ding, Y. et al., 2019).

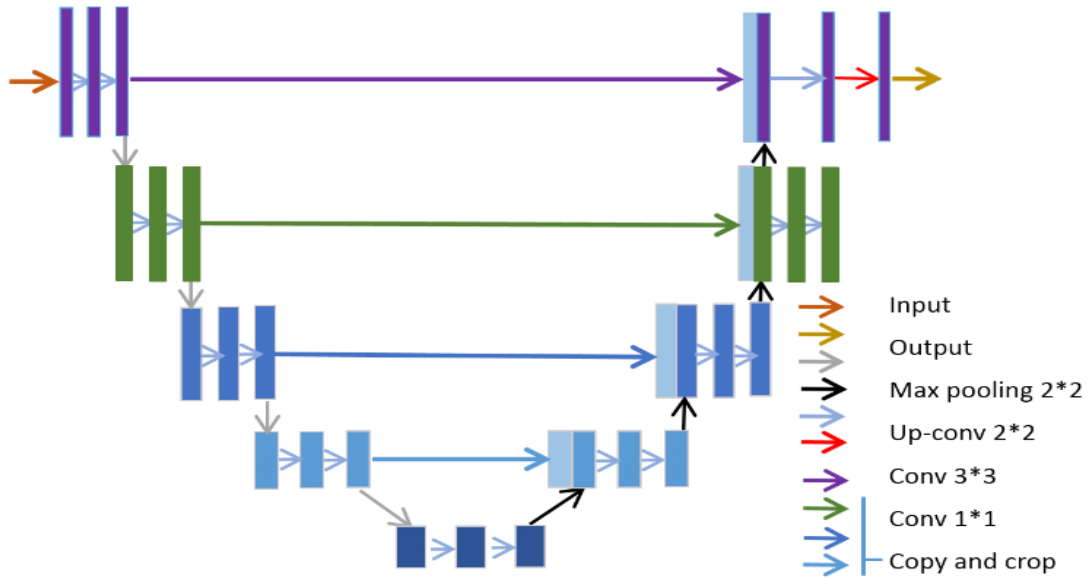


Figure 3. 6. The architecture of U-net (Ding, Y. et al., 2019).

The mathematical representation of U-Net (Krithika Alias Anbudevi & Suganthi, 2022; Tai et al., 2024) is explained below.

The encoders/contracting paths shown on the left-hand side of Figure 3.6 comprise several layers. Each layer l in the encoder is represented by the function

$$f^l = \sigma(W^l * f^{l-1} + b^l) \quad 3.22$$

where f^{l-1} represents input features from a previous layer, W^l , b^l is the convolutional weights and biases and σ represents the activation function which is ReLU in most cases.

The output of the encoder is normally max-pooled and can be represented as

$$f_{pooled}^l = MaxPool(f^l) \quad 3.23$$

The output is then passed through the bottleneck layer which is another convolution function represented by

$$f_{bottleneck} = \sigma(W_b * f^L + b_b) \quad 3.24$$

With L representing the number of layers in the encoder/contracting path. The up sampling taking place in the decoder is performed via a skip connection at every layer on the encoder. So, each layer in the decoder up samples the feature map, concatenates the corresponding encoder features and applies convolution with the given function as

$$f_{upsampled}^l = ConvTranspose(f_{bottleneck}^l) \quad 3.25$$

$$f_{concat}^l = Concat(f_{upsampled}^l, f_{encoder}^{L-1}) \quad 3.26$$

$$f^l = \sigma(W_{decoder}^l * f_{concat}^l + b_{decoder}^l) \quad 3.27$$

At the output layer, there is a 1x1 convolution layer that uses SoftMax activation to map the desired output segments and is represented below:

$$y = Softmax(W_{out} * f^{L_{decoder}} + b_{out}) \quad 3.28$$

During training of the U-Net, the loss function used is usually the cross-entropy loss and it is normally applied pixel-wise (Krithika Alias Anbudevi & Suganthi, 2022; Tai et al., 2024). Table 3.5 provides a summary of works that have used the U-Net model with medical images.

Table 3. 5. U-Net Segmentation techniques.

Author	Imaging Modality	Disease/Body Part	Variant Used
(Wang, Zekun et al., 2021)	CT	Liver/ Lung	Attention U-Net
(Khan, R. A. et al., 2022)	CT	hepatocellular carcinoma	Enhanced U-Net
(Kong et al., 2022)	CT	Liver	Enhanced U-Net
(Chetty et al., 2022)	MRI	Brain Tumour	None
(Ding, Y. et al., 2019)	MRI	Brain Tumour	Enhanced U-Net
(Huang et al., 2023)	Colour Fundus	Diabetic retinopathy	Enhanced U-Net
(Lin, Dall'ara et al., 2024)	MRI	Various/Musculoskeletal	Enhanced U-Net
(Henson et al., 2024)	MRI	Lower limb muscle	Attention U-Net/SCU-Net

(Lin, Henson et al., 2024)	MRI	Musculoskeletal	Various
(Zhou et al., 2020)	Various	Various	Enhanced U-Net (U-Net++)
(Ding, T. et al., 2024)	Ultrasound	Breast cancer	Enhanced U-Net (Attention Gate)

3.7.1 Literature review findings for the U-Net

The U-Net is mostly used for segmentation tasks, especially for segmentation of cancer for various image modalities. Over time, there have been a lot of modifications to the vanilla U-Net model as different researchers have tried to enhance the different facets of the U-Net, ranging from the skip connections to modifying the convolutional layers with attention networks in a bid to increase segmentation quality or reduce computational resource. For example, using stacked micro-CT of laminated shale, (Wang, Yongchao et al., 2023) compares random-forest segmentation with U-Net++ to map pores and microfractures, then digital-rock and Lattice Boltzmann simulations. Random forest better captured weakly connected fractures and low-intensity pixels, whereas U-Net++ missed small features. Representative-volume analysis gave porosity $\approx 6.6\%$ (RF) vs $\approx 2.6\%$ (U-Net++). Only $\sim 1\text{-}2\%$ of pores formed connected networks; microfractures ($\sim 1\%$) dominated flow. Their findings demonstrates that ML/DL segmentation enables pore-scale insight and highlight the need for vertical fracture networks to improve production.

A multi-level feature assembly MFLA U-Net was reported which is integrated with multi-scale information attention MSIA and pixel-vanishing attention mechanism by (Garbaz et al., 2024). This enhanced U-Net model was designed to boost segmentation performance. Their model was tested on different medical imaging datasets with different modalities such as colonoscopy and dermoscopic images. They have used the dice index coefficients as a metrics to evaluate the effectiveness of their developed model in segmenting these images. Their model outperformed many state-of-the-art U-Net models on the datasets used for testing. A lightweight U-Net architecture was applied on a publicly available brain tumour datasets (BraTs) to segment brain tumour (Chetty et al., 2022). The focus of the study was mainly developing a low resource U-Net framework which had a multimodal CNN encoder-decoder. They also excluded augmentation to reduce the computational demand of their network. Their model achieved a remarkable performance with dice coefficient values of up to 0.93 for specific classes they segmented when compared to other U-Net models. An enhanced U-Net model with minimal parameter was reported (Ding, Y. et al., 2019). The authors achieved this by developing a framework known as Stack Multi-Connection Simple Reducing Net, otherwise known as SRNet.

This network used fewer convolution operation in the down sampling and up sampling processes, which in turn helped to reduce the total parameter of the vanilla U-Net algorithm by 20%. They also modified the original architecture by ensuring the convolutional layers were not stacked, helping to reduce information loss. Their model was also tested with the BraTs dataset. They obtained matching results with popular variants of the U-Net model, also using the Dice coefficient parameters as an evaluation tool. The vanilla U-Net was modified for the purpose of improving the accuracy of segmentation (Khan, R. A. et al., 2022). They explored the weakness of the U-Nets models which only focus on contextual information and neglects other useful features of the channel. The developed HDA-ResUNet which combined the best of attentions mechanisms, U-Net and dilated convolution. They evaluated their model on ISI and LiTS segmentation datasets and achieved a good performance in terms of the dice coefficient. This model also used fewer parameters compared to the conventional U-Net.

3.8 Transfer learning

Deep learning techniques are known to be computationally intensive, owing to the large number of trainable parameters available in their networks. These parameters increase substantially as the network deepens. Also, the availability of large amounts of annotated medical image datasets is scarce and it is a very important element in the use of deep learning for medical image analysis (Kora et al., 2021; Yu et al., 2022). Transfer learning algorithms was developed to help solve these problems by providing a means to reparametrise an already trained large deep learning network. These networks are CNN based networks, trained on millions of images with different classes. The resulting learnt parameters are saved to be reused on other datasets. Some aspects of these networks are modified to suit the new dataset. Transfer learning involves using pre-trained deep learning models (e.g., ResNet, VGG, DenseNet, GoogleNet, XceptionNet, AlexNet, Inception V3 and SqueezeNet) and fine-tuning them on medical images. Developing models when using transfer learning is performed in two stages: Initialising the weights and finetuning. During initialisation of the weights, the weights of a previously trained model with a different dataset, i.e. (AlexNet – trained on ImageNet) as shown in Figure 3.7 (Yu et al., 2022) , are copied. When a new training dataset such as medical image in this case is used to train this model, the weights are updated. In the finetuning process, some of the CNN layers are frozen thereby the weights are not being updated. Another method of finetuning is freezing all the CNN layers barring the classification layer that is adjusted according to the requirements of the medical image data. This technique is highly effective when the available dataset is small or specific to a particular medical condition. It reduces training time and improves diagnostic accuracy (Ayana

et al., 2024). Transfer learning can be categorised into inductive, transductive and unsupervised learning based on data labels. They can also be categorised as homogenous and heterogeneous based on how consistent the dataset features and label between the source and target domain (Atasever et al., 2022; Kim, H. E. et al., 2022)

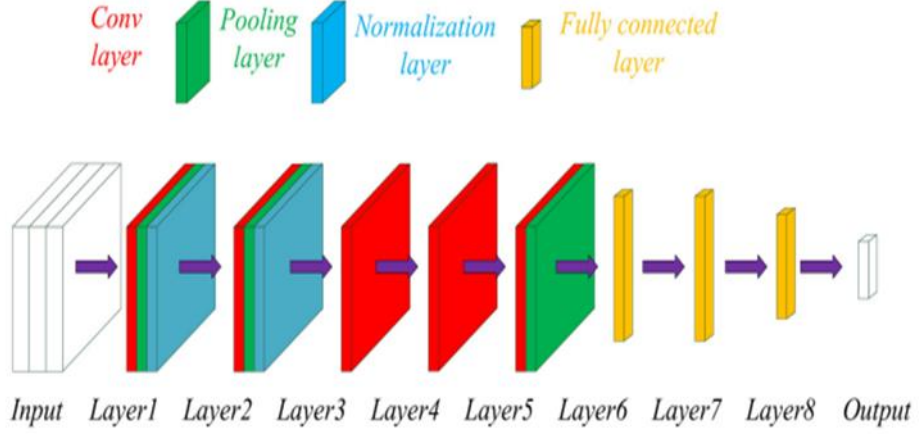


Figure 3. 7. Structure of a transfer learning model – AlexNet (Yu et al., 2022).

The concept of transfer learning can be modelled mathematically as described next (Zhuang et al., 2020). The source domain be represented as

$$D_s = \{x_s^{(i)}, y_s^{(i)}\} \text{ for } i=1, \dots, N_s \quad 3.29$$

where $x_s^{(i)} \in X_s$ is the input data from the source domain and $y_s^{(i)} \in Y_s$ represents the corresponding labels and N_s is the number of samples. The target domain be represented as

$$D_T = \{x_T^{(i)}, y_T^{(i)}\} \text{ for } i=1, \dots, N_T \quad 3.30$$

where $x_T^{(i)} \in X_T$ is the input data from the target domain and $y_s^{(i)} \in Y_s$ represents the corresponding labels N_T is the number of samples, with the assumption that $X_s \neq X_T$ or $Y_s \neq Y_T$. The objective of the transfer learning function is to find a model $f_T(x_T; \Phi_T)$ that minimises the loss in the target domain. If we represent the loss function by \mathcal{L}_T , then \mathcal{L}_T can be defined as:

$$L_T = \frac{1}{N_T} \sum_{i=1}^{N_T} l(f_T(x_T^{(i)}; \Phi_T), y_T^{(i)}) \quad 3.31$$

where l is the specific loss algorithm, which can be cross-entropy or mean squared error and Φ_T is the target domains model parameters. The transfer learning process between the source and

target domain takes two steps. We pretrain the model on the source domain, using the objective loss function given by the equation:

$$L_s = \frac{1}{N_s} \sum_{i=1}^{N_s} l(f_s(x_s^{(i)}; \phi_s), y_s^{(i)}) \quad 3.32$$

The learned parameters ϕ_s from the source domain are transferred to the target domain by freezing the base layers, where ϕ_s^{base} fixed or by fine-tuning specific layers i.e., ϕ_T^{new} .

The target domain overall model becomes:

$$f_T(x_T^{(i)}; \phi_T) = f_s^{base}(x_T; \phi_s^{base}) + f_T^{new}(x_T; \phi_T^{new}) \quad 3.33$$

The learning process is then optimised by regularising the losses. Therefore, a combined loss function L given by

$$L = L_T + \lambda \cdot L_{regularisation} \quad 3.34$$

where λ is the regularisation weight and $L_{regularisation}$ can be any term used for smooth transfer i.e., L_2 -norm (Zhuang et al., 2020).

3.8.1 Literature review findings for transfer learning

Transfer learning (TL) techniques are mostly applied for the classification of medical images as summarised by the studies listed in Table 3.6.

Table 3. 6. Summary of transfer learning techniques.

Author	Image Modality	Disease/body part	TL Variant/best Model
(Santana et al., 2018)	Histopathological Images	Breast Cancer	ResNet 50
(Kumar, S. et al., 2023)	MRI	Brain Tumour	Improved ResNet 50
(B. & Kalirajan, 2024)	MRI	Alzheimer's	Various (EfficientNet)
(Saied et al., 2023)	CT	Pulmonary Nodules	Various (DenseNet)
(Yang, D. et al., 2021)	X-ray/CT	Covid-19	Various (VGG 16)
(Odusami et al., 2021)	MRI	Alzheimer's	Modified ResNET 18
(Kalusivalingam et al., 2021)	Ultrasound	Thyroid	VGG-16

This is because the networks they are learning from have been pretrained for classification as well. Several transfer learning models such as VGG16, DenseNet 121, EfficientNet-B6 and ResNet 50 were used for the binary classification tasks of X-ray radiographs and CT images (Gupta & Sharma, 2024; Yang, D. et al., 2021). Using the FracAtlas X-ray dataset, (Gupta & Sharma, 2024) fine-tuned EfficientNet-B6 for binary fracture classification. Contrary to expectations, common preprocessing techniques such as anisotropic diffusion, Canny edges, segmentation, random central cropping) did not improve the evaluation metrics as just image normalisation alone yielded the best results. Trained for seven epochs on a Tesla P100, the model reached 96.83% test accuracy (precision 0.977, recall 0.961, F1 0.969, AUC-ROC 0.961), outperforming prior reports ($\leq 95\%$). The gains are attributed to transfer learning and dataset quality. (Kassem et al., 2023) proposed a transfer-learning model for pelvis fracture detection on X-ray, comparing AlexNet, ResNet50, and GoogleNet with Grad-CAM for visual explainability. Using 876 images (472 fracture, 404 normal), GoogleNet with transfer learning achieved the best performance of 98.5% accuracy, sensitivity, specificity, and precision, surpassing ResNet50 (94.7% across metrics) and a weak AlexNet baseline ($\approx 54\%$). Grad-CAM heatmaps highlighted clinically relevant pelvic regions, improving interpretability and user trust. A small clinical case study (15 AP films) showed GoogleNet misclassified only one image, whereas ResNet50 failed on three, supporting the model's robustness.

(Alam et al., 2025) proposed MobLG-Net, a transfer-learning feature-engineering pipeline that feeds MobileNet features to a LightGBM stage and then trains classic ML classifiers. On a 9,463-image, multi-region radiograph dataset (train 8,863/test 600), MobLG-Net features with Logistic Regression or LightGBM achieved 99% accuracy; KNN and Random Forest reached 98%. A MobileNet classifier alone achieved 96-97% accuracy, while a classical CNN using spatial features reached 81%. Computation was also improved: training time fell markedly versus spatial features (e.g., LightGBM 0.0415 s vs 29.46 s; RF 3.9 s vs 70 s).

(Guan, Q. et al., 2019) used the Inception-v3 TL model for the classification papillary thyroid carcinomas (PTCs) and benign thyroid nodules using ultrasound images. A total of 2,836 images from 2,235 patients were used. The model was trained to crop nodule margins and make diagnostic predictions. The best performance was achieved using a 50-pixel margin and 384×384 image size. In the test group, Inception-v3 achieved a sensitivity of 93.3% and a specificity of 87.4%. Their proposed model was most accurate for nodules sized 0.5-1.0 cm.

3.9 Vision transformers

Despite the significant successes recorded in the enhancement of the diagnostic accuracy of deep learning models such as CNN, RNN and U-Net in the classification and segmentation of medical images, there have remained some limitations. Their reliance on localised feature extraction, leading to inductive bias and sequential operation, make them fall short when the medical imaging tasks requiring long range dependency and global feature extraction (Tian, D. et al., 2023). Although initially designed for natural language processing tasks such as sentiment analysis (Ogunleye et al., 2024), machine translation and text summarisation (Bora & Cuayáhuil, 2024), their ability to capture long range dependencies in image pixels helps to build a more robust segmentation/classification model. Vision transformer is a relatively new deep learning architecture that is increasingly being applied to medical imaging. Developed by Google in 2020, it performs segmentation or classification using the transformer architecture. The ViT creates a partition of the input images into multiple patches of 16x16 pixels and linearly embeds them. For the pixels to be suitable for the transformer architecture, they must be transformed into fixed-length vectors (Pu et al., 2024). The self-attention mechanism represents the main feature of the ViT architecture as this forms the basis on the interaction between the pixel patches. It also uses positional encoding to represent the spatial location of the image patches. Feedforward layers, placed after the self-attention layers, are generally used in for making final decisions by the model (Berroukham et al., 2023). Several large language models have been developed for medical image analysis (Van et al., 2024); however, the focus is on the legacy transformer model for this analysis. The original architecture of the vision transformer is shown in Figure 3.8 (Dosovitskiy et al., 2021).

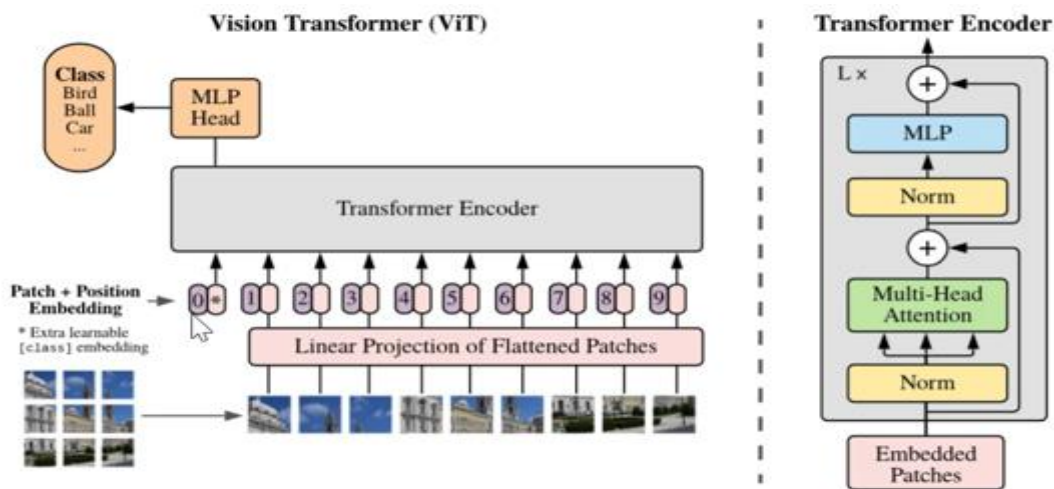


Figure 3. 8. Vision Transformer architecture (Dosovitskiy et al., 2021).

If the input to the ViT model is represented by an image X , then X can be defined such that

$$X \in \mathbb{R}^{H \times W \times C} \quad 3.35$$

Where H and W represent the height and width of the image respectively, and C is the number of channels which is usually 3 for RGB images. The input image is divided into non-overlapping patches of size P , such that

$$X_{patches} = \{x_1, x_2, \dots, x_N\} \quad 3.36$$

where $N = \frac{H \times W}{P^2}$ is the total number of patches and each patch is flattened as

$$x_1 \in \mathbb{R}^{P^2 \cdot C} \quad 3.37$$

The patches are sent through the transformer encoder layers consisting of the multi-head attention layer, which computes the relationships between patches using the K, V, Q matrices (Keys, Values and Queries), which are computed as

$$K = Z^{(l-1)}W_K, V = Z^{(l-1)}W_V, Q = Z^{(l-1)}W_Q \quad 3.38$$

where $Z^{(l-1)}$ is the positional patch embedding for each patch and W_K, W_V, W_Q is the learnable weights projection matrices, which have the same size as embedding space. This is then passed to the feed forward neural network (FFN) and a classification head for a classification task (Han et al., 2023).

3.9.1 Literature review findings for vision transformers

Several studies used ViT and its variants including TNT, Swin, DeiT and PVT (Han et al., 2023) for classification, registration and segmentation of medical images. Just like the convolution-based transfer learning models, learnable parameters from transformer models such the ones listed above can also be used for specific DL tasks. For example, a pretrained Swin transformer was used for the classification of breast cancer using publicly available breast X-ray images (Tanimola et al., 2024). The dataset was resized to fit the Swin transformer input size and augmentation was also performed to improve generalisation. They achieved very high classification rates based on their selected metrics for evaluation with an accuracy of 99.9% and a precision of 99.8%. They compared their results with convolution-based TL algorithms (ResNet50 and VGG16) and found that the Swin transformer had a superior performance.

On 4,207 annotated proximal femur radiographs, a Vision Transformer (ViT) model was proposed by (Tanzi et al., 2022) to classify AO/OTA subtypes with 83% test accuracy (macro

precision 0.77, recall 0.76, F1 0.77), outperforming an InceptionV3 CNN (67%) and a hierarchical CNN cascade (61%). Attention maps concentrated on clinically relevant regions. ViT embeddings showed superior unsupervised separability (clustering accuracy 0.78; NMI 0.61; ARI 0.72) versus CNN features (0.36; 0.22; 0.22) and image autoencoder baselines (0.22; 0.01; 0.01). In a controlled reader study (11 clinicians, 150 images), ViT alone achieved 0.91 accuracy, while CAD assistance raised clinicians' accuracy from 0.68 to 0.97. (Chład & Ogiela, 2023) evaluated vision transformers for cervical spine fracture detection on CT slices and outline. They used a YOLOv5 detector to crop the vertebrae, used windowed slices for patch-embedding, and trained ViT variants with augmentation. They also incorporated explainability by attention rollout. The smallest DeiT-T16 attained the best accuracy (98%) after 200 epochs, outperforming ViT-B16 (91% with augmentation) and ViT-B32 (92%); without augmentation ViT-B16 reached 85%. YOLOv5 achieved $mAP \approx 0.98$ for vertebra localisation. Their results indicate ViTs are competitive with CNNs for slice-wise fracture detection and practical for real-world triage when paired with robust augmentation and cropping (Abbaoui et al., 2024). (Guan, B. et al., 2024) proposed a one-stage hybrid detector for thighbone fracture localisation that couples an enhanced Pyramid Vision Transformer (with overlapping patch embedding) to an FPN neck and a new spatial- and scale-aware attention (SSA) block. On a curated, radiologist-annotated dataset of 4,000 thighbone X-rays (train 3,600/test 400), the model achieves AP 53.7% and AP50 87.0%, surpassing recent CNN/Transformer baselines finetuned under identical settings. Ablations show the improved PVT backbone contributes +4.5 AP/+4.3 AP50, while the SSA block adds +5.3 AP/+4.8 AP50; using both yields the best performance. Transformers were used for image registration tasks (Chen, J. et al., 2022). In their study, they developed the TransMorph algorithm which is a hybrid transformer and convolution network. The network leveraged on the encoder/decoder architecture of transformers but instead of the attention mechanisms at the decoder, this was replaced with a convolution network. For the transformer network, they used the Swin variant due to its ability to extract feature maps at different resolutions by merging patch layers, making it suitable for the image registration task. The algorithm was tested on different image pair datasets comprising mainly MRI and CT modalities for registration purposes. They obtained very competitive results based on the Dice score evaluation when compared to both traditional and other DL methods used for similar tasks. (Wang, Chuantao et al., 2022) proposed the use of ViT model with deep neural network called the ConvTrans-Net for the segmentation of lymph node tumours using Ultrasound images from Kaggle. The model works by concatenating the different feature vectors of the ultrasound images and passed on to a Multilayer Perceptron and output of the multiple attention mechanism is passed to a feed-forward layer for

segmentation purposes. Their model was evaluated using the Jaccard similarity coefficient, precision and recall, and was benchmarked with the NCP model, and WS-2017 models. They obtained a Jaccard coefficient 85.21%, recall of 89.65% and a precision of 85.17, when compared to the benchmark models.

3.10 Hybrid models

Most deployed deep learning methods use CNN as described in the preceding sections, with variants such as VGG, ResNet, LSTM, RNN, GAN and GRU all being used for different medical image analysis. However, different authors have tried to combine the strength of some of the models to deal with the weaknesses of the others by combining them together and this forms the basis for hybrid models (Zhu, Zirui, 2024). Some works have also combined different deep learning methodologies by focusing on the strengths of the methodologies. For example, a convolutional network's strong local feature extraction has been combined with the long-range dependencies of transformers when performing medical analysis.

3.10.1 Convolutional-based hybrid models

A hybrid model of convolution algorithms was developed comprising SegNet, MultiResU-Net and Krill Herd Optimisation algorithm (KHO) to improve the segmentation of CT scans of liver lesion and RNA genome sequencing (Ramamurthy et al., 2020). The SegNet framework provided the segmentation capacity of their model, utilizing pixel-wise classification through the Softmax layer. A CNN based architecture was employed with the MultiResUNet to handle the lesion segmentation, together with the SegNet framework. The hyperparameters of the models, when optimised through the KHO algorithm, helped to improve the segmentation process. They tested their algorithm with a publicly available LiTs datasets and used evaluation metrics such as the Dice coefficient, F1-score and accuracy and obtained better results with F1-score comparatively higher than the models used for comparison.

(Yadav et al., 2022) proposed a two-scale Hybrid SFNet which is a hybrid model of CNN and Canny edge algorithm, that fuses gray-level X-rays with improved Canny-edge maps to better localise fracture boundaries before classification. On an augmented, class-balanced dataset of 34,000 images drawn from public sources and split 80/20, SFNet achieved the best results among tested models. Reported performance reached 99.12% accuracy with $F1 \approx 0.99$; per-class metrics in Table 4 show fracture precision 1.00 and recall 0.98 (healthy precision 0.98, recall 1.00), exceeding AlexNet/VGG16/ResNeXt (91–95%) and far surpassing MobileNetV2 (48%). Confusion matrices highlight SFNet's fewest errors (0 false positives, 60 false negatives) versus large false-positive/negative counts for MobileNetV2.

A hybrid model of convolutional methods involving ResNet and U-Net model (ResU-Net) was developed for the segmentation of liver and tumours using CT images (Rahman et al., 2022). Their ResU-Net model focused on improving the available models by providing improved image contrast and segmenting irregular tumour shapes and small tumour sizes. The combination uses the best of ResNet's residual connection and the U-Net's encoder and decoder structure to enhance feature learning, segmentation precision and efficiency. They also implemented various augmentation techniques such as rotation and reflection to increase the variability in their dataset. They achieved an accuracy of 99.6% and a dice coefficient of 99.2%. (Zhu, Yangyang et al., 2022) developed a hybrid model for the classification of unexplained cervical lymphadenopathy (CLA), using ultrasound images of patients in underdeveloped area of China. The CLA-HDM model was made up of three smaller models, each designed to handle a specific diagnosis task related to unexplained lymph node abnormalities (CLA). Model 1 checked whether the issue was benign (non-cancerous) or malignant (cancerous). Model 2 looked deeper into benign cases to decide between tuberculosis or a reactive condition. Model 3 examined malignant cases to determine if it was due to metastasis (spread from another cancer) or lymphoma. Each model had two input branches, one for grayscale ultrasound (BUS) images and another for colour Doppler (CDFI) images. The CDFI images were processed to emphasise important colour details. Then, both images were analysed by a deep learning model (ResNet-50). The efficacy of their model was evaluated using Area under curve which was above the 0.8 benchmark for each model. The authors in (Vasile et al., 2021) developed an ensemble model of a custom-built convolutional neural network (5-CNN) and a transfer learning model using the pre-trained VGG-19 architecture for the classification of thyroid disorders using ultrasound images of four categories: autoimmune, nodular, micro-nodular, and normal. The combined CNN-VGG method showed superior performance, achieving a test accuracy of 97.35%, specificity of 98.43%, and sensitivity of 95.75%. It also demonstrated strong predictive capabilities, with high positive and negative predictive values, and an area under the ROC curve of 0.96.

3.10.2 Convolution-transformer based hybrid models

For fracture classification, (Khan, M. I. et al., 2025) proposed a hybrid Vision Transformer-Convolutional Neural Network (ViT-CNN) for elbow fracture detection on X-ray, fusing CNN local features with ViT global context via feature-fusion before classification. Using the MURA elbow subset and extensive augmentation, the dataset was expanded to 8,000 images and split 80/20 for training/testing. Their model achieved 99.93% accuracy, 99.88% precision, 100% recall, 99.94% F1-score, and an AUC of 1.0. The confusion matrix reported 805 true positives,

794 true negatives, one false positive, and zero false negatives. Comparative analyses indicated clear gains over prior methods (e.g., hybrid DL 84.99% accuracy; VGG-16+ViT 82.88%; Inception-ResNet-V2 79.6%). A hybrid model called TBConvl-NET that combines CNN, LSTM and ViT for the segmentation of several diseases was developed using publicly available datasets of different modalities such as ultrasound and MRI (Iqbal et al., 2025). The hybrid model targeted some well-known challenges in segmenting medical images such as scale, texture and shape of pathology. Due to high computational resources required, they used a depth wise separable convolution, thereby reducing computational overhead. Swin transformer blocks were used in the skip connections to help deal with the varying scales of the data and help preserve semantic information. They used Dice index, accuracy and Jaccard index to evaluate their model and compared the developed model with other hybrid segmentation models and obtained improved results. A hybrid classification model that combines the transformer and the convolution model to improve the classification of skin lesion was developed using publicly available datasets (Obayya et al., 2023). They used the Swin-U-Net architecture to perform image segmentation, leveraging self-attention of the Swin transformer and robust hierarchical analysis of the U-Net. They combined this with the Xception and ResNet 18 models for feature extraction to further improve on the image analysis. For hyperparameter tuning, a Hybrid Salp Swam Algorithm (HSSA) was used to obtain the optimal parameters, hence avoiding local minima during training. A gated recurrent unit (GRU) network was used for the eventual classification. They achieved an accuracy of 94.51% and 95.38% on both datasets used in the study. Their model performed better when compared to TL models like AlexNet and ResNet18. (Yan et al., 2024), proposed a framework that focuses on the three-vessel view (3VV) that identifies three major heart vessels: the pulmonary artery, aorta, and superior vena cava using ultrasound images. In the first step, a YOLOv5 (based on CNN architecture) model was used to detect these vessels and earmark the region of interest. In the second step, they used a modified DeepLabv3 model with a new Attentional Multi-scale Feature Fusion (AMFF) module to perform segmentation. Using a dataset of 511 images, the model achieved high accuracy, with Dice scores of over 85% for the pulmonary artery and aorta.

3. 11 Discussion

This review provided a broader view of various AI and DL methodologies used for medical imaging analysis. These included Convolutional Neural Networks (CNNs), Adaptive Neuro-fuzzy Inference Systems (ANFIS), Recurrent Neural Networks (RNNs), Autoencoders,

Generative Adversarial Networks (GANs), U-Net architectures, Vision Transformers, and hybrid models. For each of these methodologies, some of the key findings were presented.

CNNs are extensively used for medical imaging tasks like disease detection, classification, and segmentation and the most important contribution of the algorithm is the unrivalled feature extraction through convolutional layers, making them effective for tasks requiring spatial hierarchy. CNN models can be lightweighted in terms of computing resources as inspired by ResNet while also balancing the model performance. Deeper CNN architectures can be computationally expensive but can be very useful in multi-disease detection, for example in the classification and detection of COVID-19 and viral pneumonia, achieving high accuracy due to advanced hardware. They can also be effective when the task includes a pixel-level feature extraction. Examples can be seen in its use in the ultrasound image classification of fatty liver. RNNs, particularly their variants like Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), are suitable for sequential data tasks. Their ability to retain temporal dependencies is utilised in dynamic imaging and disease progression monitoring. RNNs which usually have convolution layers are effective for classification tasks, with different imaging modalities, with MRI dominating. They were applied to detecting several diseases such as Alzheimer's disease and breast cancer. LSTM variants of the RNNs algorithm handle tasks like segmentation of temporal MRI images and noise removal in diagnostic images. Some modifications have also been applied to the RNN algorithm training, such as using the Honey Badger Algorithm (HBA) for optimisation of the model training parameters.

Autoencoders are unsupervised models used for dimensionality reduction, anomaly detection, and data augmentation. Their encoder-decoder structure helps compress and reconstruct medical images with minimal distortion. Autoencoders are employed for image analysis tasks like segmentation and denoising and have been used with several image modalities for the analysis of different diseases. Enhancing autoencoders by fusing with wavelet transform can also help improve noise removal from medical images. They can also be combined with convolutional networks to improve disease classification. GANs are used for generating synthetic medical images, data augmentation, and improving image resolution or quality. The main components of a GAN network comprise: a generator for creating synthetic images and a discriminator for distinguishing between real and generated images. GANs can be very computationally intensive due to the dual deep convolutional networks. Research around this area is focussed on reducing models' parameters and maintaining performance. Different variant of GANs have employed different image analysis tasks with each variant tuned to fit the specific tasks.

U-Nets are specialised for medical image segmentation and have been widely adopted for tasks like tumour detection and organ delineation. The architecture combines encoder-decoder pathways with skip connections for improved feature representation. Different variants of the U-Net architecture also exist to enhance the vanilla model, making them suitable for specific image modality while performing its main task of segmentation. Also, enhancements have been created to reduce model operating costs, while retaining the segmentation process. More recent studies have been adding transformers to U-Nets to further enhance segmentation accuracy. Transfer learning leverages pretrained models (e.g., ResNet, VGG, DenseNet) on large datasets and fine-tunes them for medical imaging tasks. They are mostly used for classification tasks. Models like VGG16 and ResNet50 perform well for COVID-19 classification, achieving very high accuracy on different image modalities. Some transfer learning models are quite deep with millions of parameters, making them very computationally intensive. Custom hybrid models pretrained on unlabelled medical datasets have also been used to improve classification accuracy for skin lesions and breast cancer images, mimicking the transfer learning ideology. They provide an option in terms of computing resources and domain specific training and learning. ViTs are a relatively newer deep learning architecture, adapted for medical imaging tasks. They require global feature extraction and long-range dependencies. They divide images into patches and process them using self-attention mechanisms. ViT models can also be pre-trained like most CNN models and the learning transfer for a newer task. Several variants have also been developed depending on the task. With larger training datasets, ViTs have shown to outperform CNN models when used for tasks with fewer annotated data.

Hybrid models combine the strengths of multiple architectures (e.g., CNNs with RNNs, or CNNs with transformers) to improve medical image analysis. CNN-based hybrids like ResUNet combine ResNet's feature extraction with U-Net's segmentation capabilities. They were used widely for cancer related image segmentation. Convo-transformer hybrids, like TBConv1-NET integrate CNN, LSTM, and ViT for improved segmentation and classification of diseases. Hybrid models can also be optimised. The algorithm such as hybrid salp swarm algorithm (HSSA) could be used to obtain the optimal parameters thereby increasing model performance.

3.12 Summary

This review explored the transformative impact of artificial neural network techniques, including deep learning algorithms and transformers architectures on medical image analysis. CNNs are widely used for feature extraction and disease classification across medical imaging modalities, achieving high accuracy with optimised computational costs. ANFIS and some of its

modifications are widely used for classification tasks. Recurrent neural networks (RNNs), including LSTM and GRU, enhance temporal analysis for disease progression monitoring. Autoencoders and generative adversarial networks (GANs) assist in data augmentation, denoising, and synthetic image generation. U-Net architectures improve segmentation for tumour detection and organ delineation. Vision transformers (ViTs) leverage attention mechanisms for superior classification and registration. Hybrid models combining CNNs, transformers, and optimisation techniques enhance performance, while transfer learning mitigates data scarcity, ensuring robust results across imaging applications. Together, these advancements underscore the versatility and efficiency of deep learning in medical diagnostics, paving the way for improved clinical outcomes and personalised healthcare solutions. Future advancements may focus on computational efficiency, integrating multimodal data, and enhancing interpretability for clinical adoption.

Chapter 4 – Methodology

In this chapter, the general methodology which is related to the obtained results is presented. To ensure clarity and ease of reading, the research methodologies used for specific studies are presented and explained in greater details in their respective chapters (Chapters 5-8). This chapter is divided into two main sections: Data collection and classification pipelines. The first section discusses the process of recruiting participants and recording the IRT images used for this study. The second section briefly describes the data analysis pipelines including MLP, ANFIS and CNN techniques used for the IRT image classification.

4.1 Data collection settings

Participants for this study were recruited within the Emergency Department of Sheffield Children Hospital. Suitable conditions for collecting the thermographic imaging were ensured by selecting a room close to X-ray radiography laboratory to minimise patient's stay at the hospital. The selected room was deemed the most suitable as it is devoid of any heat sources from the external environment that could affect the accuracy of the thermal images. The room had a window, which was not opened throughout the imaging collection. The room temperature was monitored every time a patient was brought in for thermography and a consistent value ranging from 18-25 °C was measured which is the recommended for temperature values of a recording room for thermography according to (Diakides et al., 2012; Ring & Ammer, 2012). The collected images were then taken to Sheffield Hallam University, where analysis was performed in MATLAB, Excel and Python.

4.2 Study ethics approval

The related ethical process was considered during the course of data collection by applying for ethical approval from the NHS Research Ethics Committee (REC) on 7th of March 2019 (IRAS project ID: 253940). Sheffield Hallam University ethics approval was also gained during the course of the study.

4.3 Other processes of ethical considerations

Throughout the duration of this research, we ensured that utmost care was taken to ensure that other areas of research ethics are strictly adhered to and met with professionalism. The ethical considerations were outlined in the next sections.

4.3.1 Informed consent

As the study recruited children, their parents or legal guardian give consent on their behalf and informed consent was ensured. To obtain the informed consent, the parent or legal guardian of the participants of the research were provided with an information sheet. For children aged above 3 years, an age-appropriate participation sheet was also provided to them (see Appendix i - iv). The researcher also provided a brief verbal overview of the study and allowed the parents or legal guardian of the participants an ample time to make decisions on participation without having any influence on their final decision whether to participate or not.

Following this process an agreement to participate in the research was achieved by assenting (for children) or consenting (for the carers or parents) (Appendices v and vi). The researcher taking the consent during the research process had completed the relevant approved e-learning ethical considerations for courses such as Ethical concerns associated with different form of research and activities, and working ethically in challenging circumstances among others (Appendix vii).

4.3.2 Welfare of participants and adverse event

During the course of this study, it was ensured any risk to participants were minimised. The thermal camera used for the recording of the thermal images was non-invasive and there was not any adverse effect recorded with the equipment during the study. The duration of the data recording was within 10 seconds.

4.3.3 Confidentiality of patients

Only the authorised medical staff for the participants had access to confidential information of the patients, including patient's notes. An authorised clinician at the hospital anonymised the names and other related data for transferring them to the University. At all times, it was ensured that stringent care was taken when handling this information from patients and utmost confidentiality of the patients whose information we have was maintained.

4.3.4 Data storage and protection

The collection and storage of data for the study duration were carried out in accordance with the Data Protection Act (1998) following the appropriate guidance and consent. The collected data, consisting of thermographic images and Excel files that contains related anonymised patients data were transferred to Sheffield Hallam University research folder (Q drive) which is a very secure outlet and is only accessible by the research team members at the University.

Anonymity of patient data was ensured by coding sensitive information within the spreadsheet and a metadata for discerning the codes was created in a document we referred to as Patient Recruitment Log. Other than the infrared thermal images, the patient's sex, date of birth, time admitted, time left the hospital, injury outcome confirmed by an X-ray (fracture or sprain), type of fracture, time since injury, location of fracture and any medication taken were recorded. A snippet of the Excel file showing some of the anonymised patient information is provided in (Appendix ix).

4.4 Recruitment process

Patients scheduled for a diagnostic wrist X-ray between March and June 2019 at the Sheffield Children's Hospital were considered for inclusion in the study. Participation was only offered to those who were already set to undergo an X-ray, ensuring that any part of the study did not alter any aspect of the patient's diagnostic process or treatment plan.

4.4.1 Target group

The study aimed to enrol an equal number of male and female participants aged 5 to 15 years who were undergoing a diagnostic wrist X-ray at the Sheffield Children's Hospital Emergency Department.

4.4.2 Inclusion and exclusion criteria

To ensure informed consent was properly obtained, exclusion criteria included cases where the patient or caregiver might struggle to understand the study, such as non-native English speakers or individuals with cognitive impairments. Other exclusions included children with multiple injuries or wrist deformities, those experiencing severe pain, or patients categorised above triage category D. Furthermore, any patient who appeared highly anxious or had not provided consent or assent was also excluded.

4.4.3 Recruitment procedure

Upon arrival at the ED, triage records were reviewed to determine whether a patient met the eligibility criteria. Those considered suitable, were initially approached in the x-ray waiting area, where both the child and their parent/legal guardian received a brief explanation of the study and were given corresponding information sheets. They were then given time to review the documents and consider participation (only children above 3 years were given assent and their information sheet).

4.5 Data collection process

The data collection was assisted by a medical student at the time, Ms Charlotte Reed.

4.5.1 Key considerations in data collection development

4.5.1.1 Video recording vs. still images

Before commencing data collection, it was determined that video recording would be more effective than still images for capturing skin surface temperature as it allowed averaging of the resulting frames to reduce thermal noise. This method accounted for small temperature fluctuations, thereby improving accuracy of the measurement. A 10-second thermal video of the wrist placed in a flat position on a table was used in the study. Frame capture rate was 30 frames per second thus producing 300 frames. Video recording allowed averaging of images thus reducing noise as compared to a single image recording.

4.5.1.2 Tracking algorithm

A tracking algorithm was employed to align images making up the video prior to averaging. Tracking was needed as during recording there could have been small hand movements. In the case where the movement was too large and the algorithm failed to track the image correctly, such images were excluded. For this study, the Kanade–Lucas–Tomasi (KLT) gradient-based tracker was utilised (Fassold et al., 2009; Usman et al., 2019). The KLT tracker is widely used because it is computationally efficient for point tracking and has shown strong performance relative to competing approaches (Usman et al., 2019). Before tracking begins, feature locations are defined within a region of interest (ROI). These features are image points that are visually distinctive from their surroundings, such as L-shaped corners, T-junctions, or isolated bright spots on a dark background. The KLT algorithm first identifies such unique feature points within the ROI and then attempts to follow those same points across subsequent frames (Fassold et al., 2009).

KLT operates through three main components. First, a *feature detector* selects salient points in each frame that are most likely to remain identifiable in the next frame (based on large eigen values of the image pixel gradient); the output is an array of pixel coordinates. Second, a *feature tracker* takes these coordinates and estimates where each feature has moved in the following images. Some features inevitably drop out as tracking continues, for example if the point leaves the camera's field of view, becomes occluded, or changes appearance so much that it cannot be reliably matched. Third, a *feature manager* maintains adequate coverage and quality of the tracked set. When features are lost, the manager triggers the detector to add new points to replace them, while also promoting an even spatial distribution of tracked features across the image. This

helps preserve robustness, improves distinctiveness, and prevents the tracker from relying too heavily on a small or clustered subset of points (Fassold et al., 2009; Usman et al., 2019).

For successful tracking to occur, multiple boxes segmenting the ROI due to movement of the hand are seen according to the diagram in the Figure 4.1.

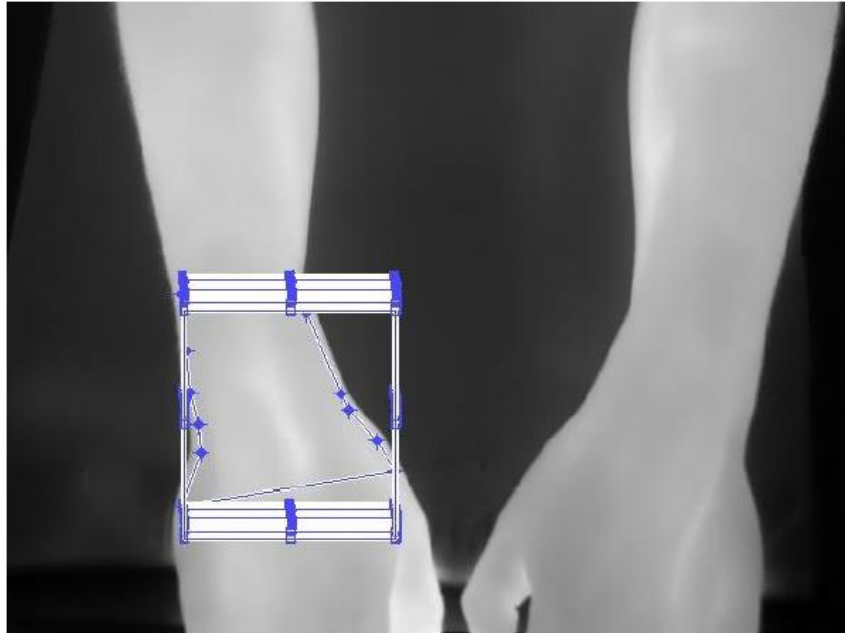


Figure 4. 1. Depicting the successful tracking of the ROI (Reed et al., 2020).

4.5.2 Data collection protocol

Once informed consent/assent was obtained, the child was comfortably seated at a table where the infrared camera, securely mounted on a tripod, was positioned. For imaging, the camera was carefully adjusted so that its lens remained perpendicular to the child's arms. This setup was achieved with the use of a boom arm and counterweight to maintain stability and precision. Additionally, to bring the arms closer to the camera for optimal imaging, they were positioned on a box covered with an insulating material, as illustrated in Figure 4.2.

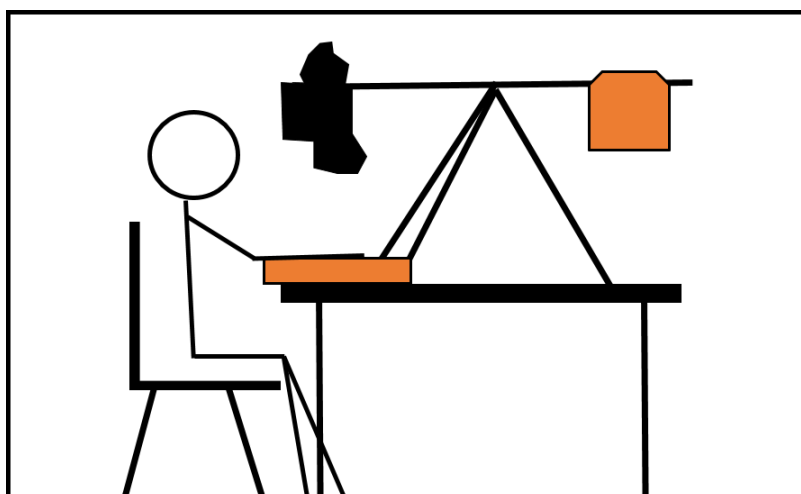


Figure 4. 2. A schematic representation of the initial data collection protocol (Reed et al., 2020).

Before the imaging process, the camera was powered on for a few minutes to calibrate the electronics. To reduce heat transfer between the insulating material and the patient's arms, patients were instructed to position their arms on the insulating material just before recording began. Whenever possible, their arms were kept apart and fully pronated. Once correctly positioned, patients were asked to remain still, and the infrared thermal video was recorded. This approach ensured minimal interference from body heat and maintained consistency in the imaging process.

Once imaging was completed, the patient's body temperature was measured using a tympanic thermometer. Both the patient and their guardian were given the opportunity to ask any additional questions. They were then provided with the study questionnaire (Appendix viii) and are requested to return to the x-ray waiting area, where they could complete the questionnaire while awaiting their x-ray results.

4.6 Research equipment, software and image processing

4.6.1 Thermal camera and software

The thermal imaging in this study was conducted using the FLIR T630sc handheld camera, which was securely mounted on a tripod for stability. The detailed specifications of this camera are provided in (Appendix x). To operate the camera and capture thermal images efficiently, a laptop was used, running the FLIR Research Max 4 software. This software facilitated image recording and analysis, ensuring precise data collection. A screenshot illustrating the interface of the FLIR Research Max 4 software is included in Figure 4.3 (Reed et al., 2020), providing a visual representation of the tools used for processing and analysing the thermal imaging data.

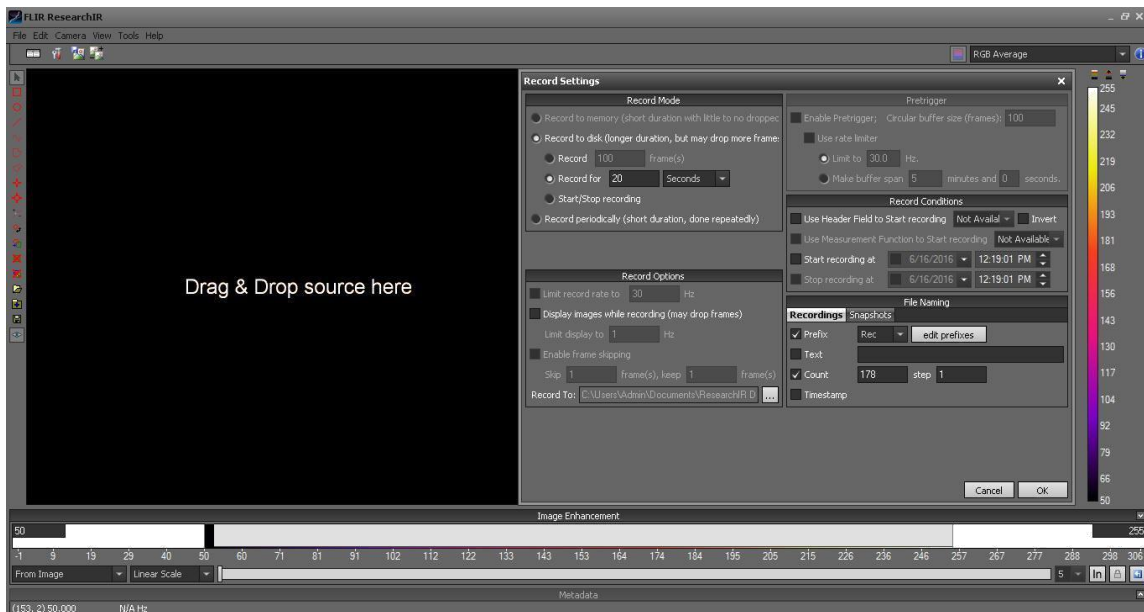


Figure 4. 3. Showing the screenshot of the GUI of the FLIR Research Max 4 software that controls the FLIR T630sc camera (Reed et al., 2020).

4.6.2 Analysis software

Information related to skin surface temperature and machine learning diagnosis was processed using Excel and MATLAB[®]. Python programming was also used during later stages of development.

4.6.3 Selecting the ROI and image processing

For the wrist region, the ROI was defined to include the carpal bones along with the distal radius and ulna. The selection process was guided by anatomical landmarks visible in each image. Initially, an area was chosen that closely aligned with the space between the radiocarpal joint and the base of the metacarpal bones as shown in the figure 4.4 (Reed et al., 2020). The distal wrist joint line (radiocarpal joint) is used as the primary reference point, identified between the carpal bones and the distal radius/ulna. Proximal and distal boundaries are established using fixed proportional distances along the forearm, ensuring consistent ROI extraction across subjects. The coloured arrows represent measured segments that standardise the ROI height and position relative to anatomical structures. This approach minimises variability and ensures that the selected ROI consistently captures clinically relevant regions associated with wrist pathology for subsequent analysis. The length of this region was then used to establish a proportional measurement ratio, which was applied to determine the final ROI dimensions.

This standardised approach ensured that despite differences in patient anatomy, the selected wrist and arm regions were consistently measured for all participants. Figure 4.4 (Reed et al., 2020) illustrates the ratio-based measurement method used for ROI selection, demonstrating how each region was systematically identified and processed for analysis (Reed et al., 2020).

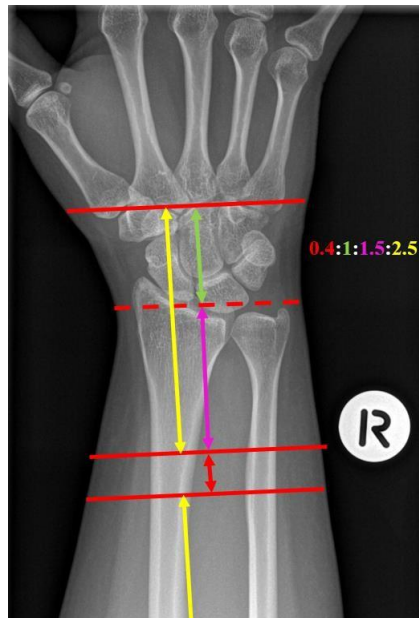


Figure 4. 4. Showing the diagrammatic representation of the ratios used in determining the ROIs for the wrist and arm regions (Reed et al., 2020).

Each selected ROI contained multiple pixels, with each pixel representing a specific temperature value. To determine the average temperature for each ROI, the temperatures of all pixels within the region were averaged. This process was carried out for all 300 frames in the recorded image, and a final average temperature was calculated across all images. Based on the patient's diagnosis, either a fracture or a sprain; each individual had nine average ROI temperature measurements. This method ensured consistency in temperature data collection and allowed for accurate comparisons between different injury types, providing a reliable dataset for evaluating thermal variations in injured and non-injured wrists and arms.

4.7 Machine learning and deep learning analysis.

In this study, the ML and DL techniques used includes Multilayer perceptron, Convolutional neural network and Adaptive Neuro fuzzy inference systems. The methodological pipeline used in this study is presented in their appropriate chapters. Here, a brief of ML analysis used in this work such as feature extraction, cross validation and image augmentation is briefly presented.

4.7.1 Feature extraction

Feature extraction is used to transform the raw IRT images into compact, discriminative representations that enhance class separability while suppressing nuisance variability which could be noise, pose, or illumination. In this study, 3 main feature extraction techniques were applied to the IRT images obtained from the participants. This includes statistical measures, Grid search and Frequency domain representation. All these feature extraction techniques are manually applied to the IRT image before the neural network and ANFIS processing.

4.7.1.1 Statistical representation of the ROI

For the MLP machine learning pipeline, the ROI of the participants wrist IRT features were represented with nine (9) main statistical measures of central tendencies and dispersion namely: maximum, minimum, mean, standard deviation, median, mode, skewness, kurtosis and interquartile range. The same was performed for the uninjured wrist, used as reference. These features were used for the MLP and ANFIS pipelines presented.

4.7.1.2 Grid search

To enhance better discrimination between the temperature difference of injured and uninjured hand feature extraction process, the averaged injured-wrist ROI was partitioned into 10×10-pixel cells in form a rectangular grid; each cell was represented by its mean temperature. The 50 highest cell means were selected as candidate “hot” areas, then normalised by subtracting the contralateral ROI mean to mitigate inter-subject variability. From these 50 values, nine statistics were computed including max, min, mean, median, mode, standard deviation (SD), skewness, kurtosis, and interquartile range (IQR). Based on discriminative analysis, three features namely: standard deviation, kurtosis, and interquartile range were retained as inputs.

4.7.1.3 Frequency domain representation

The obtained IRT images were also represented by their frequency domain representation. Frequency-domain representations describe images in terms of spatial frequencies i.e., how rapidly intensity changes across space rather than raw pixel values (Gonzalez, 2009). The 2-D Discrete Fourier Transform (DFT) was used in this study to decomposes the IRT image into sinusoidal basis functions. The magnitude spectrum encodes dominant scales and orientations (e.g., periodic textures), while the phase preserves geometric structure (edges, contours). This feature extraction technique was used for the CNN pipeline presented in chapter 7 to help enhance discrimination of the IRT images as they are visually similar.

4.7.2 Normalisation

Normalisation process was used to place features of the IRT images on comparable scales to improve numerical conditioning, stabilise gradients, and make distance- or margin-based learners used for this study behave sensibly (Bishop, 1994). For the MLP and ANFIS pipeline, min-max scaler was used to scale the feature to a value between 0 and 1.

4.7.3 Cross validation

Cross validation technique was used in this work due to the number of datasets available for the classification algorithms. The small amount of dataset hinders the selection of training, testing and validation percentage of the datasets in order not to overfit or underfit learning networks. Cross validation can be computationally intensive; but for the purpose of this work, accuracy of the classifications outweighs the computational cost of the model and since the datasets are small, it is highly recommended. K-fold cross validation was implemented in this study for building the ANFIS pipeline. This is done because there are many parameters used in fine-tuning the ANFIS (see chapter eight).

4.7.4 Image augmentation

Image augmentation is the artificial way of expanding the training datasets when constructing a ML or DL pipeline. This involves the applying label-preserving transformations to the existing images in order to help improve generalisation and reduce overfitting especially when dataset is limited (Shorten & Khoshgoftaar, 2019). Classic geometric (flip, rotate, scale, crop, translate) and photometric (brightness/contrast, noise, blur) transforms promote invariance to pose, illumination, and image acquisition variability. For this study, image augmentation was performed on the IRT images for the CNN pipeline (see chapter seven) after the FFT feature representation. The augmentation techniques used rotation, translation, shearing and noise.

4.8 Summary

This chapter provides the methodology used in the paediatric wrist injury assessment with infrared thermography (IRT) for this study. Data were collected in a controlled room adjacent to radiography room at Sheffield Children's Hospital; ethics approvals from NHS REC and Sheffield Hallam University were obtained, with parental consent/child assent, anonymisation, and secure storage (Q drive). Ten-second, 30fps IRT videos were recorded with a FLIR T630sc; frames were aligned via tracking and averaged to reduce noise. Standardised ROI definitions covered carpal bones and distal radius/ulna on both wrists; average temperatures per ROI were derived across 300 frames and paired against the contralateral side. Analysis used MATLAB,

Excel, and Python. Machine-learning pipelines included MLP, CNN, and ANFIS. Feature extraction comprised of nine statistical representations, a grid-based “hot-spot” search with three selected statistics (SD, kurtosis, IQR), and Fourier-magnitude features for CNNs. Min–max normalisation, k-fold cross-validation, and post-FFT augmentation (rotation, translation, shear, noise) supported training and evaluation.

Chapter 5 – Correlation of Skin Temperature with Time Since Injury in Paediatric Wrist Injuries: An Infrared Thermal Image Analysis

5.1 Introduction

The purpose of this study reported in this chapter is to use infrared thermal imaging to improve the understanding of temperature changes taking place during fracture healing. The findings could then improve means of detecting fractures using infrared thermal imaging as well as overall understanding of fracture healing.

A number of studies used infrared thermal (IRT) imaging to assess skin temperature changes at bone fracture sites. Infrared (wavelength 700 nm to 1 mm) is part of the electromagnetic spectrum and is emitted by objects with a temperature above absolute zero (-273.15 °C or 0 °K). The mean temperature difference of a fractured distal arm in 25 patients (mean age 65.9±10.4 years) was compared to their healthy arm (Haluzan et al., 2015) using IRT imaging. The study showed an increase in temperature difference of 1.20 °C and 1.42 °C after 1 and 3 weeks respectively and a decline thereafter. IRT imaging of forearm fractures and the contralateral (uninjured) side in 19 children (aged 4 to 14 years) indicated an overall temperature difference of 0.13 °C after a day, 1.17 °C after 1 weeks, 0.83 °C after 2 weeks, 0.23 °C after 3 weeks and 0.14 °C after a month (Ćurković et al., 2015). IRT imaging of 11 children aged 5-18 years diagnosed with osteogenesis imperfecta (a disorder resulting in the bones to be more fragile) could differentiate between fractured and healthy thoracic vertebrae (Fane De Salis et al., 2018). IRT images of forty children (mean age 10.5 years), 19 with wrist fracture and 21 with wrist sprain was studied (Reed et al., 2020). It was reported that fractured wrists had an overall significant temperature increase of 1.52% compared with the uninjured contralateral wrists. Although sprained wrists also had an increase in temperature when compared to the uninjured wrists, their temperature increase was not statistically significant. A study of IRT imaging in 113 children (aged 1 to 14 years) with traumatic injuries indicated the method can be valuable for ruling out fractures (Sanchis-Sánchez et al., 2015).

In this chapter, an exploration of the correlation between skin temperature at the injury site (wrist fracture or sprain) and time since injury (TSI) in paediatric and the manner the relationship differs in the two types of injuries is presented. In the following sections the methodology and materials are explained, results are presented and discussed, and the study's conclusion is provided.

5.2 Chapter related methods

In this section, the methodology used to obtain results included in this chapter are explained.

5.2.1 Participants' details

The study included forty participants, 24 males and 16 females, mean age 10.50 years (standard deviation 2.63 years), 19 with a wrist fracture and 21 with a sprain. The diagnosis was confirmed by an X-ray radiograph. Thirty participants had analgesic medication, mainly paracetamol and ibuprofen. Mean body temperature was 36.3°C (standard deviation 0.43°C) across all participants.

5.2.2 Image recording

The participants waited in the recording room for 10 minutes with both wrists uncovered, to allow acclimatisation to the room temperature. The time since injury measurement was taken from each participant as recorded in the patient anonymised data provided in the Appendix ix. For recording, both hands were placed on a mat covering a table in front of them. The mat insulated the hands from temperature influence of the table. The recording room temperature was within the recommended 18-25°C which is acceptable for human IRT recording (Ring & Ammer, 2012). To reduce background temperature affects, draught and external heat sources in the recording room were minimised. The IRT image recordings were carried out using a FLIR T630sc handheld camera. The camera was positioned 1 metre above the wrists. The camera's specifications are noise equivalent temperature difference (i.e., minimum temperature difference resolvable by the IR camera) less than 30 mK, image resolution 640×480 pixels, spectral range 7.5 µm to 13 µm, dynamic range 14 bits and operating temperature -40 °C to 650 °C. To facilitate initial data storage, the camera was connected to a computer using a USB cable. Image capture rate was set to 30 frames per second (i.e., the camera's maximum capture rate for 640×480 pixels resolution), emissivity was set to 0.97 (this is suitable for recording from human skin). A recording duration was 10-seconds resulting in 300 images. The resulting images were averaged to reduce thermal noise, producing a single image.

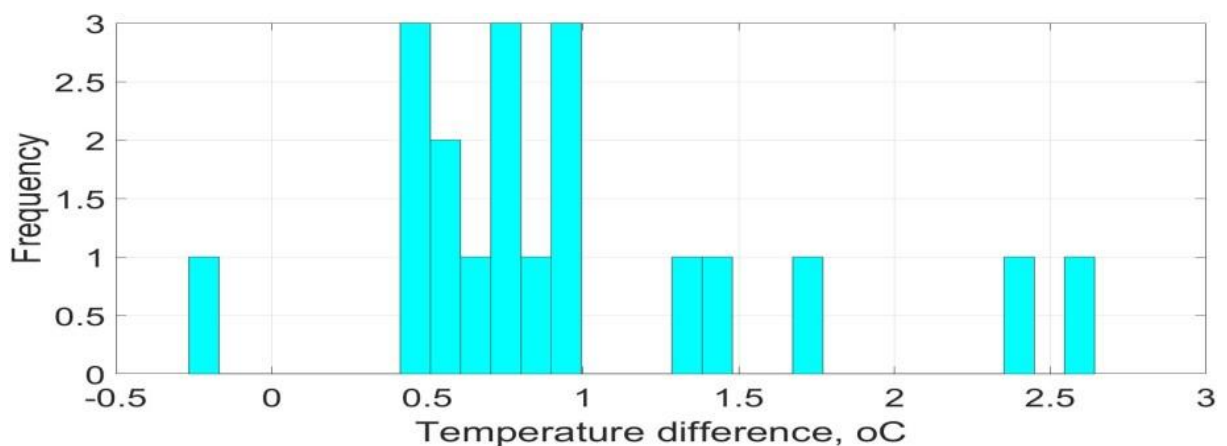
5.2.3 Region of interest (ROI) segmentation

Image processing was performed using Matlab[®] package. The first image of the recorded image was displayed and a region that included the carpal bones and a section of the distal radius and ulna was cropped. Template matching (Fassold et al., 2009; Usman et al., 2019); Lewis, 1995; Munsayac et al., 2017) was then used (using the cropped ROI as the template) to extract and align the corresponding region from the remaining 299 images. The extracted ROIs from the images were then averaged to reduce thermal noise. This resulted in a single image for feature extraction. ROI Representation: For the contralateral (uninjured wrist), the background section of the averaged ROI (cropped from the original image) was excluded by thresholding. The background

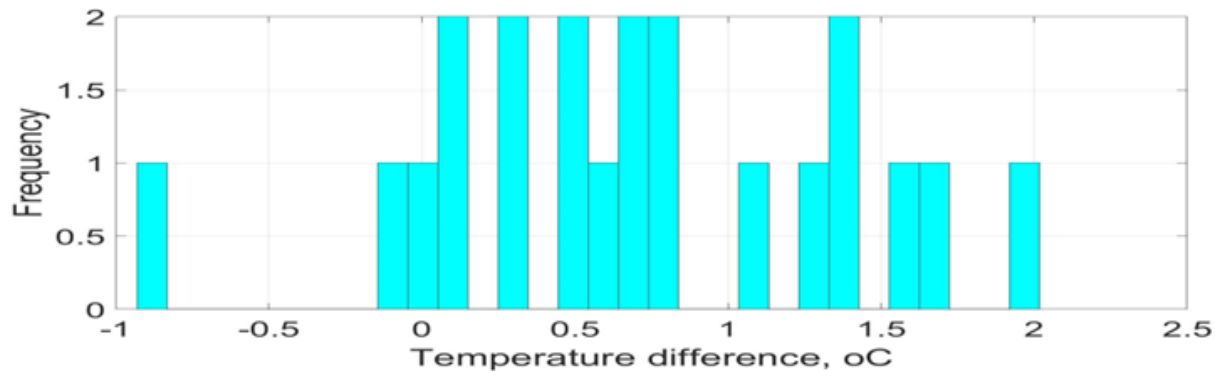
following the cropping process were all zeros and so the threshold level was set accordingly. The remaining pixel values represented the ROI temperature and were averaged to obtain an overall reference temperature. For the injured wrist (fractured or sprained), the averaged ROI was converted to a grid of cells with dimension 10×10 pixels. Each cell was expressed by its mean value. The 50 largest average values were selected, and the mean temperature of the uninjured contralateral wrist was subtracted. Mean cell temperatures were sorted in descending order, and experiments determined that selecting the top 50 values optimised MLP discrimination. Using more values introduced irrelevant regions, while fewer risked missing the injury site. The contralateral wrist mean was subtracted to account for temperature variation. Statistical features, including mean, standard deviation, skewness, kurtosis, and IQR, were then extracted. The subtraction of the contralateral ROI temperature was carried out to deal with variations of skin temperature across the participants. The resulting values were then represented by their overall mean temperature difference value (ΔT).

5.3 Results

The histograms of ΔT values for fracture and sprain are shown in Figure 5.1(a) and figure 5.1(b) respectively. For fractures, most ΔT values were between 0.5 to 1 °C. For sprain, ΔT values were more uniformly spread across their range. These indicate that fracture tends, in general, cause a more consistent skin temperature increase at the site of injury than sprain. This may be because of the relatively fixed amount of energy required to cause a fracture and hence the amount of damage and resultant inflammatory changes is more homogenous. Because sprains are more variable in terms of the trauma required to cause them, the resulting inflammation, the IRT change is more heterogenous.



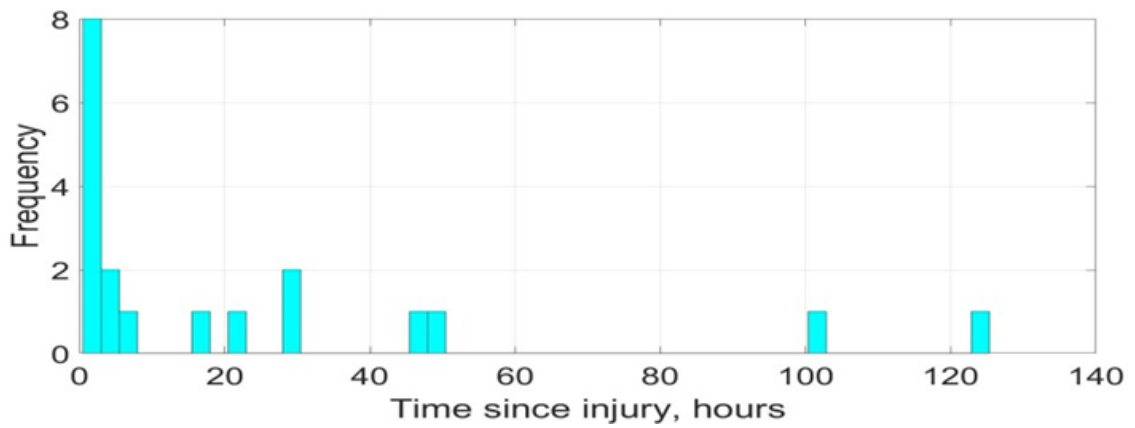
(a)



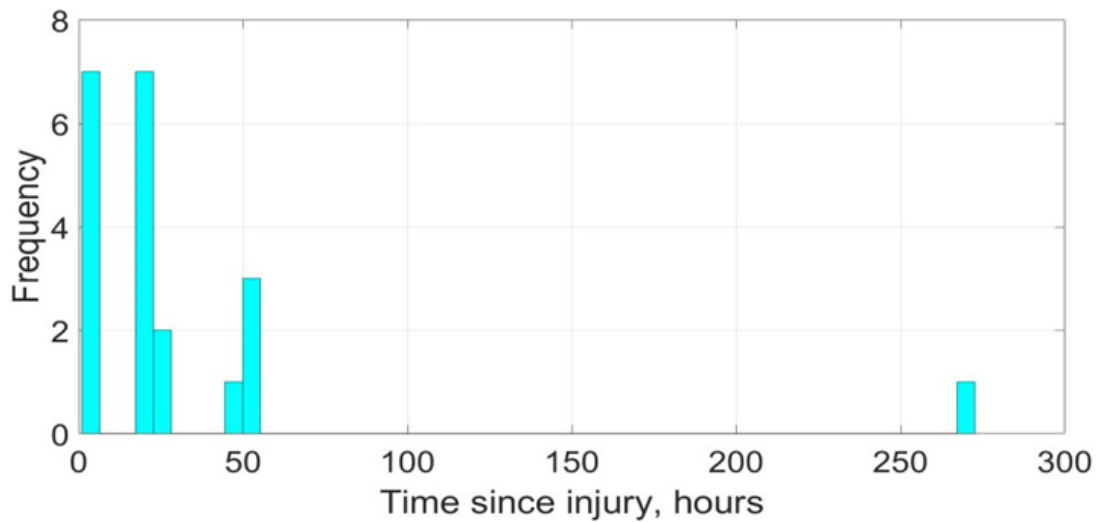
(b)

Figure 5. 1. Histograms of ΔT for (a) fracture, (b) sprain.

Figures 5.2(a) and 5.2(b) shows the TSI histograms for fracture and sprain respectively.



(a)



(b)

Figure 5. 2. Histograms of time since injury for (a) fracture, (b) sprain.

As indicated in Table 5.1, the mean TSI for fracture and sprain were 23.7 and 32.8 hours respectively.

Table 5. 1. Mean and standard deviation of time since injury for fracture and sprain.

Injury type	Time Since Injury	
	Mean	Standard deviation
Fracture	1419 (minutes)	2130 (minutes)
	23.7 (hours)	35.5 (hours)
Sprain	1966 (minutes)	3450 (minutes)
	32.8 (hours)	57.5 (hours)

As indicated in Table 5.2, the mean of ΔT for fracture and sprain were 0.962 and 0.707 °C respectively. Therefore, fracture had a higher mean ΔT a lower standard deviation for ΔT than sprain.

Table 5. 2. Mean and standard deviation of temperature difference representing the ROI (ΔT).

Injury type	ROI temperature representation (ΔT , °C)	
	Mean	Standard deviation
Fracture	0.962	0.690
Sprain	0.707	0.699

Figure 5.3 shows the plot of ΔT against the TSI for wrist fracture. Two distinct phases were observed.

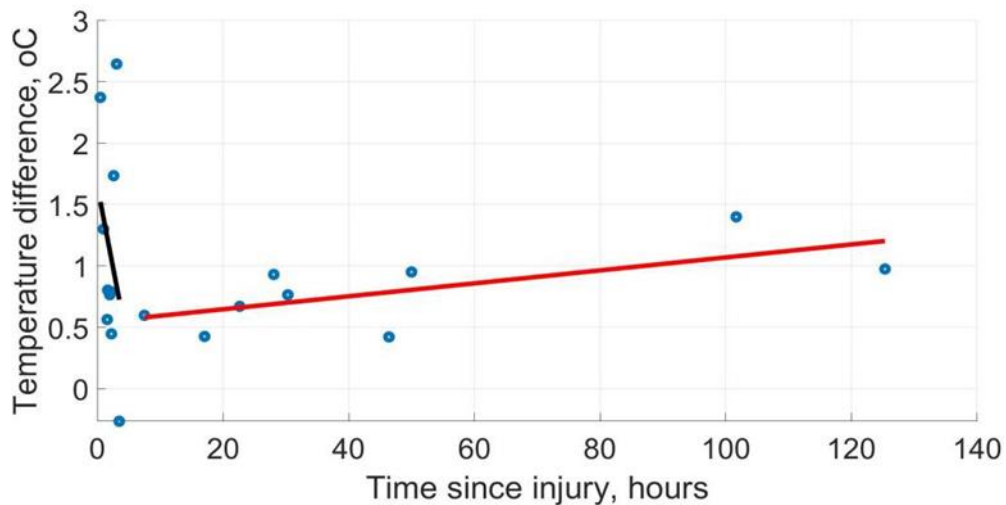


Figure 5. 3. Plot of ΔT against TSI for fracture.

The first phase, lasting up to around 3.3 hours, ΔT and TSI were not significantly correlated (correlation coefficient= -0.267). This part is separately plotted in Figure 5.4 for further clarity.

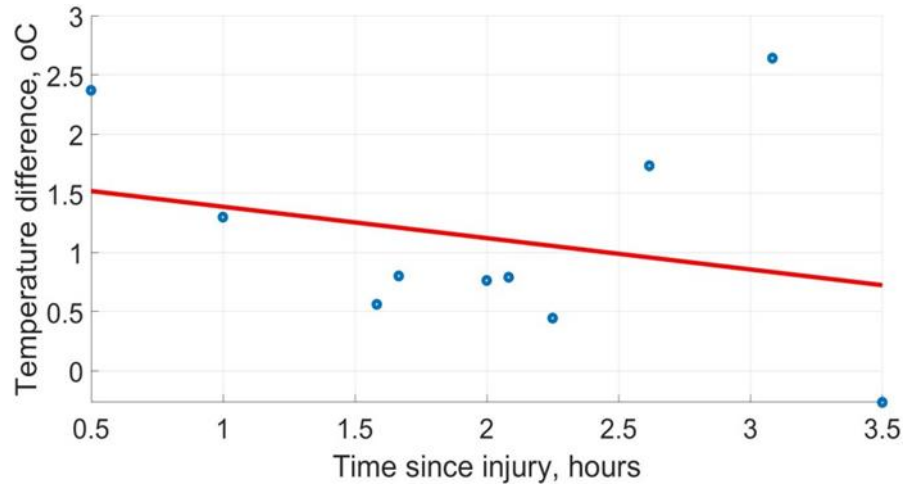


Figure 5. 4. First phase of plot in Figure (5.3).

For clarity, the second phase is shown separately in Figure 5.5. This phase was from the end of the first phase to the time of last data recording, i.e., from 3.3 hours to 125.3 hour. For this phase, a moderate correlation existed between ΔT and TSI (correlation coefficient = 0.680). For the full period of data recordings (i.e., the two parts combined), there was not a significant correlation between ΔT and TSI (correlation coefficient= -0.009).

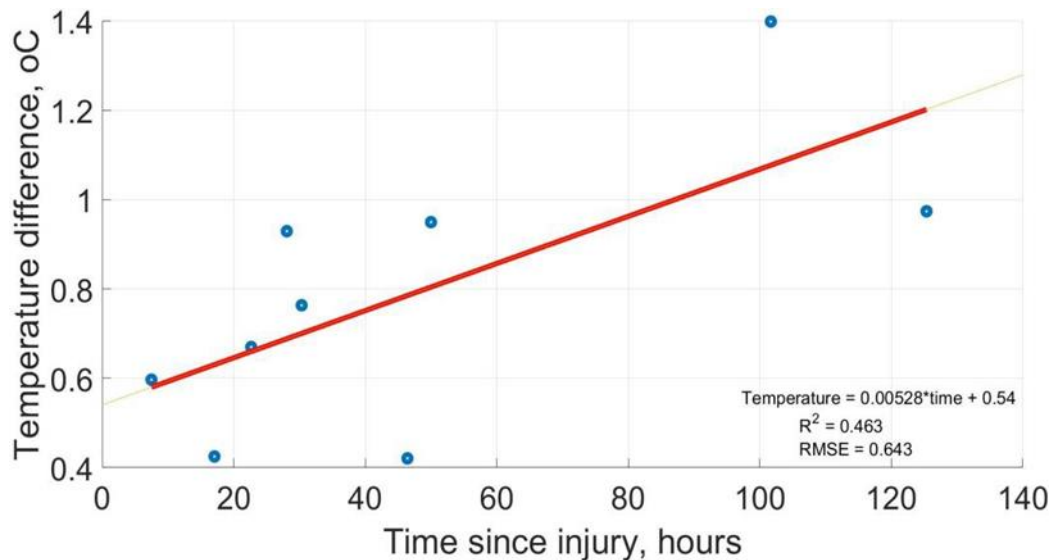


Figure 5. 5. The second phase of plot in Figure (5.3).

Figure 5.6 shows plot of ΔT against TSI for sprain. For sprain, unlike fracture, distinct parts were not observed and the correlation between ΔT and TSI for the full recording duration was not significant (correlation coefficient = -0.245).

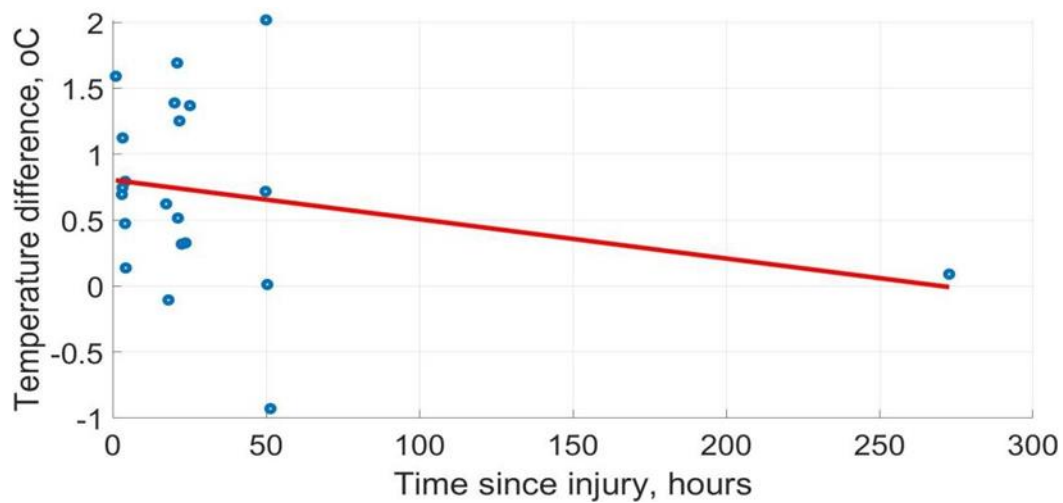


Figure 5. 6. Plot of ΔT against TSI for sprain

5.4 Summary

This pilot study explored the manner in which the skin temperature at the site of wrist fractures and sprains correlated with the time since injury (TSI). The physiological differences between the healing processes of fractures and sprains are reflected in the manner skin temperature increased at the injury site. For fracture, there was no significant correlation between temperature and TSI initially but thereafter a moderate correlation was observed. No significant correlation was observed between these variables in sprain. The findings highlighted the way in which fractures and sprains affect the temperature surrounding tissues at the injury site and the information attained may be valuable in better understanding the healing processes in the two types of injuries.

The theoretical foundations presented in Chapter 2 provide an essential framework for interpreting the findings reported in this chapter. Chapter 2 explains the physiological mechanisms underlying fracture and sprain injuries, highlighting the role of inflammation, increased blood perfusion, and metabolic activity in fracture healing. These processes result in localised elevation in skin temperature at the injury site, whereas sprains typically exhibit less consistent thermal responses due to differences in tissue damage and healing dynamics.

Chapter 2 also introduces infrared thermography (IRT) as a non-invasive modality capable of capturing temperature variations through emitted infrared radiation. The theoretical premise that

temperature asymmetry between injured and contralateral regions can serve as a diagnostic indicator underpins the analytical approach used in the analysis.

The results in also aligns with these theoretical expectations, demonstrating that fractured wrists show a measurable increase in temperature relative to the uninjured side, consistent with inflammatory and vascular responses. Furthermore, the observed relationship between temperature and time since injury (TSI) in fractures reflects the staged healing process. In contrast, the absence of a strong temperature (TSI relationship in sprains supports the theoretical distinction between bone and soft-tissue injury responses). It may also be valuable for the ongoing research in the use of IRT imaging to screen for bone fractures.

Chapter 6 – Infrared Thermal Imaging and Artificial Neural Networks to Screen for Wrist Fractures in Paediatrics

6.1 Introduction

The purpose of the study in this chapter is to observe how AI models such as Multilayer perceptron artificial neural network model could aid in classifying fracture and sprain IRT images using the manual statistical feature representations of the IRT images which includes the standard deviation, kurtosis and interquartile range.

A significant proportion of X-ray radiographs for wrist fracture diagnosis fail to demonstrate a fracture. A study of 1223 children with wrist trauma reported that 51% had a wrist fracture and the rest had normal radiographs (Slaar et al., 2012). A tool which allows for rapid, non-irradiating and easy to use screening to assist clinicians in deciding which patients require X-ray radiography could be beneficial in reducing the number of unnecessary X-rays, time spent in the emergency department (ED) and associated costs.

Infrared (IR) thermal imaging (IRTI) is a well-established technology for condition monitoring in the industry; however, its applications for medical diagnosis and monitoring are currently evolving (Owen, Ruaridh & Ramlakhan, 2017). IR is part of the electromagnetic spectrum, radiated from objects with a temperature above absolute zero, i.e., $-273.15\text{ }^{\circ}\text{C}$ (0 Kelvin). Its harmless nature and ability to indicate temperatures very accurately in a non-contact manner have made it a technology of growing interest for medical diagnosis and monitoring (Usamentiaga et al., 2014). There is evidence of an increase in temperature of the surrounding tissues at the site of the bone fracture due to increases in metabolism and blood flow (Haluzan et al., 2015). This temperature increase can be quantified, analysed, and interpreted through IR thermal image processing.

6.2 Chapter related methodology

6.2.1 Evaluation statistics

This section briefly outlines the statistical measures used to analyse effectiveness of the IRTI method in differentiating between wrist fracture and sprain. These measures are illustrated in Figure 6.1, and further related information can be found in (Lalkhen & McCluskey, 2008; Ramlakhan, Shammi L. et al., 2022). In the analysis, X-ray radiography was used as the gold

standard, as this is what is used in clinical practice. Figure 6.1 (Shobayo et al., 2022) shows the statistical measures used to evaluate the effectiveness of the MLP model.

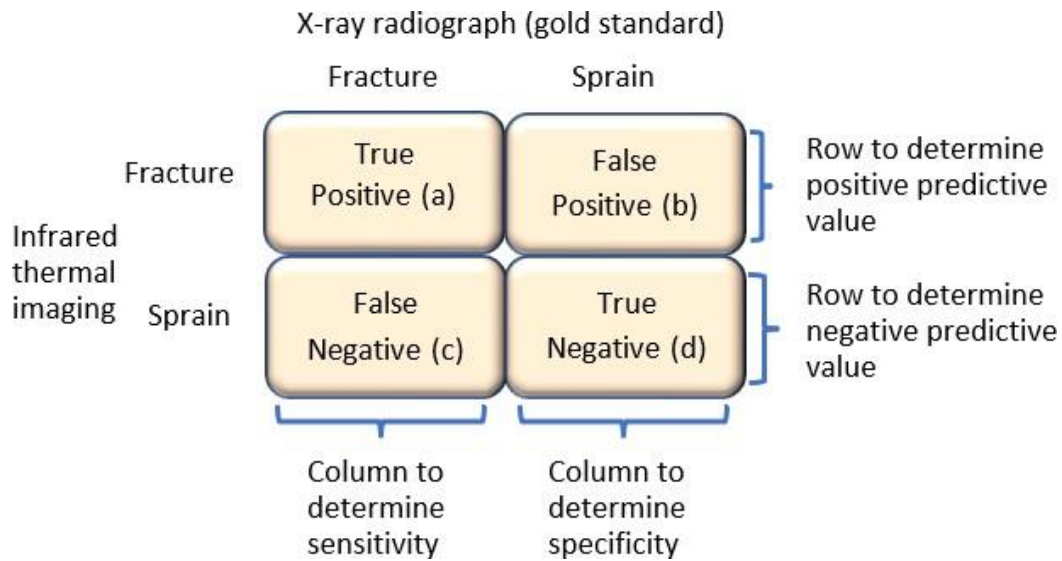


Figure 6. 1. Statistical measures used to analyse effectiveness of MLP and IRTI to differentiate wrist fracture and wrist sprain (Shobayo et al., 2022)

The measures were:

- True positives, TP, (*a*): number of participants with wrist fracture (confirmed by x-ray) correctly identified as fracture by IRTI.
- False positives, FP, (*b*): number of participants with wrist sprain (not-fracture, confirmed by x-ray) misidentified as fracture by IRTI.
- False negatives, FN, (*c*): number of participants with wrist fracture misidentified as sprain by IRTI.
- True negatives, TN, (*d*): number of participants with wrist sprain correctly identified as sprain by IRTI.
- Sensitivity: the percentage of true positives (fractures) correctly identified by IRTI, i.e.,

$$Sensitivity = \left(\frac{a}{a + c} \right) \times 100 \quad (6.1)$$

- Specificity: the percentage of the true negatives (sprains) correctly identified by IRTI, i.e.,

$$Specificity = \left(\frac{d}{b + d} \right) \times 100 \quad (6.2)$$

- Positive predictive value: IRTI-identified percentage of participants with positive results (identified as fracture) who have fracture, i.e.,

$$Positive\ predictive\ value = \left(\frac{a}{a + b}\right) \times 100 \quad (6.3)$$

- Negative predictive value: IRTI-identified percentage of participants with a negative result (identified as sprain) who do not have fracture, i.e.,

$$Negative\ predictive\ value = \left(\frac{d}{c + d}\right) \times 100 \quad (6.4)$$

- Accuracy: IRTI-identified proportion of true results, either true positive or true negative, in a population. It measures the degree of veracity of IRTI as the fracture screening scheme (Shobayo et al., 2022).

$$Accuracy = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) \times 100 \quad (6.5)$$

6.2.2 Selection of region of interest and tracking algorithm

MATLAB[®] package was used to display the first image of the recorded video. Using its cropping function, a region that included the carpal bones and a section of the distal radius and ulna was selected. This region is indicated by dotted lines in Figures 6.2 and 6.3 for fractured and sprained left wrists, respectively. In Figure 6.2 (a) and Figure 6.2 (b), the left wrist region appears brighter (indicating increased temperature) as compared to the contralateral (uninjured) right wrist.

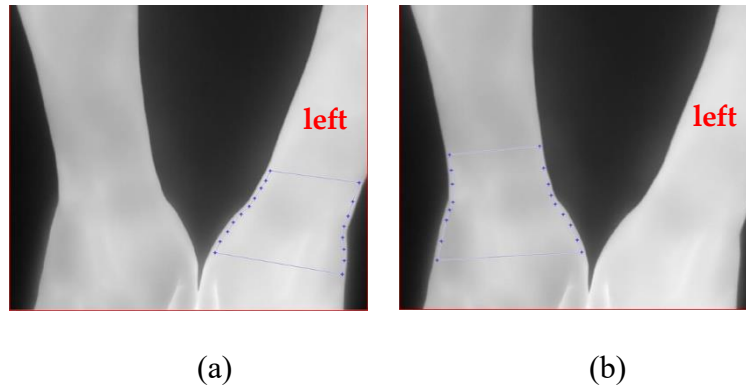


Figure 6. 2. IR thermal images indicating the region of interest as a dotted line: **(a)** fractured left wrist; **(b)** uninjured right wrist (Shobayo et al., 2022)

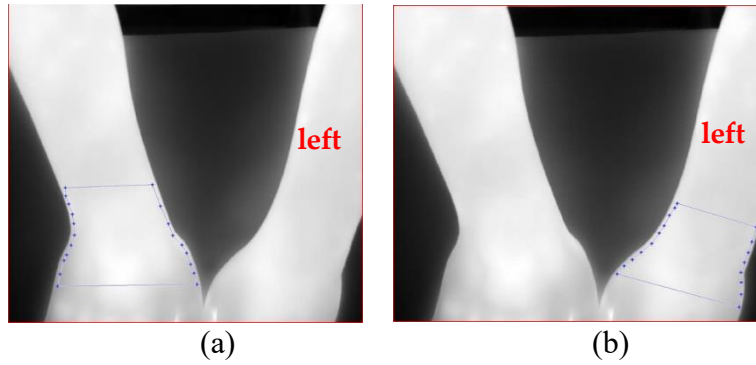


Figure 6. 3. IR thermal images indicating the region of interest as a dotted line: **(a)** sprained left wrist; **(b)** uninjured right wrist (Shobayo et al., 2022).

Any hand movements during the recording misaligned the ROI selected from the first image and the corresponding region in the following 299 images. A template matching tracking method was applied to realign the ROI across all images. Template matching measures the similarity between two images based on their normalised cross correlation (Lewis, 1995) and has been reported to be more accurate than methods such as the sum of absolute difference and sum of squared difference (Munsayac et al., 2017).

Once the ROIs from all images in a recording were extracted and aligned, they were averaged to produce a single ROI image. This process was repeated for both the injured and uninjured wrists. Figure 7.4 shows a typical averaged ROI. Averaging of the images enhanced them by reducing the effect of thermal noise.

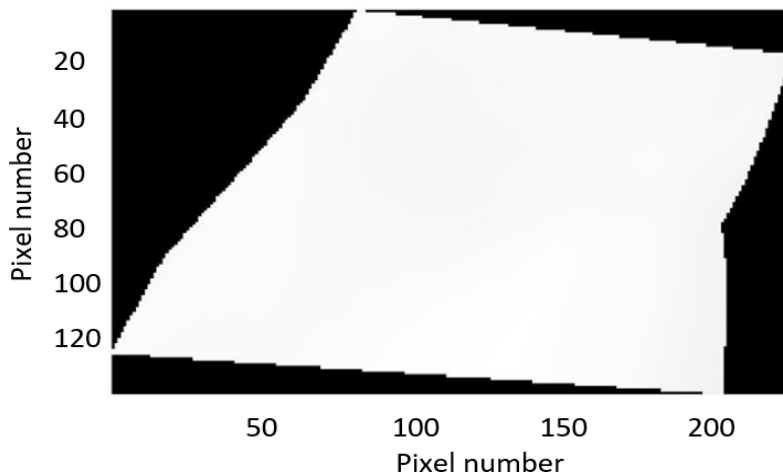


Figure 6. 4. A typical averaged region of interest (Shobayo et al., 2022)

6.2.3 ROI feature extraction

The averaged ROI needed expressing by representative features for analysis by the MLP neural network. For the contralateral (uninjured wrist acting as reference temperature), initially, the background section of the ROI (shown black in Figure 6.4) was excluded through thresholding. The background region had zero values, and the threshold level was set accordingly. The remaining pixel values representing the wrist temperature were averaged to determine an overall reference temperature.

For the injured wrist (fractured and sprained), initially, the averaged ROI was converted to a grid, consisting of cells of 10×10 pixels. The dimension of the cells was a compromise between coarseness, representing a larger region by each cell and finer resolution allowing for a greater spatial characterisation. Each cell was then represented by its mean temperature. Figure 6.5 shows the ROI represented in grid form.

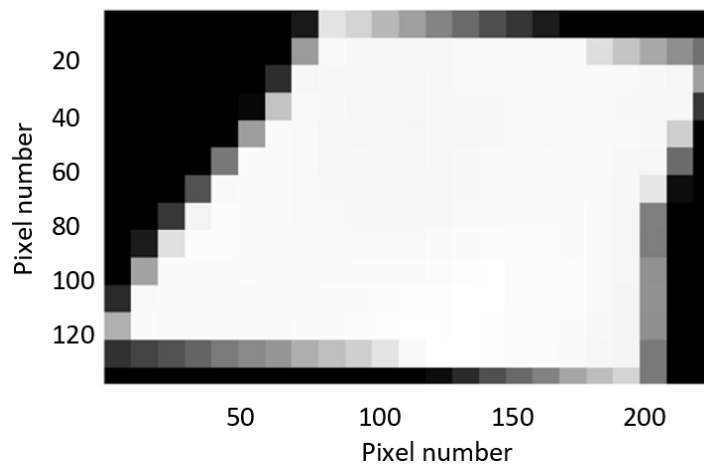


Figure 6. 5. The region of interest of figure converted to cells of 10×10 pixels, and each cell represented by its mean temperature value (Shobayo et al., 2022).

The cells with greater relative temperatures were more likely to be associated with the injury location within the segmented wrist ROI. The mean temperature values for the cells were sorted in descending order of magnitude. Experiments were carried out by selecting different numbers of largest temperature values and observing the MLP discrimination results. This led to the selection of the 50 largest values (the total number of cells within an ROI varied depending on the participant as the wrist sizes varied). Selection of more than 50 values meant broadening the area within the ROI for input to the MLP, resulting in inclusion of possible areas not associated with the injury site. A smaller number of values could have caused the injury site being

inadequately represented. The mean temperature representing the contralateral wrist was subtracted from the selected 50 values. This was carried out to deal with the skin temperature variations across the participants. The statistics of the resulting 50 values were obtained by considering their maximum, minimum, mean, standard deviation, mode, skewness, kurtosis, and interquartile range (IQR). The above operations were repeated for the 40 participants.

The justification for converting the injured wrist ROI into a grid form was that an injured wrist ROI had multiple areas with distinctly higher temperatures from its remaining parts. Initially, attempts were made to locate the injury site through clustering of the ROI; however, the approach was not effective, as multiple clusters formed, and it was unclear which cluster represented the injury site. The grid structure and selection of 50 highest temperature values ensured that the injury site was included in the analysis while the relatively cooler areas (not associated with the injury) were excluded. The approach of averaging the pixel values across the whole ROI for the injured wrist was also considered and proven not as effective, as it diminished the effect of temperature increase at the injury site by including pixels not in the vicinity of the injury site.

6.2.4 Discrimination using multilayer perceptron neural network

MLP is a well-established artificial neural network capable of pattern recognition (Bishop, 1994; Popescu et al., 2009; Ramlakhan, Shammi et al., 2022). It does not assume its input data to be from a particular type of distribution or to be linearly separable. It consists of interconnected processing elements (also known as neurons) arranged as input, hidden and output layers. An MLP (shown in Figure 6.6) with a single hidden layer was used to differentiate between fracture and sprain based on the features extracted from the IRTI.

It is important to note that time since injury (TSI) was not included as an input variable in the MLP model. This decision was based on the findings presented in Chapter 5, where TSI demonstrated limited relevance in distinguishing between fracture and sprain injuries within the studied sample. Specifically, the analysis showed no meaningful correlation between TSI and the observed temperature difference (ΔT) for sprain cases, and only a weak or inconsistent relationship for fracture cases in the early stages. As a result, including TSI as a feature would not have contributed significantly to improving the discriminative performance of the model.

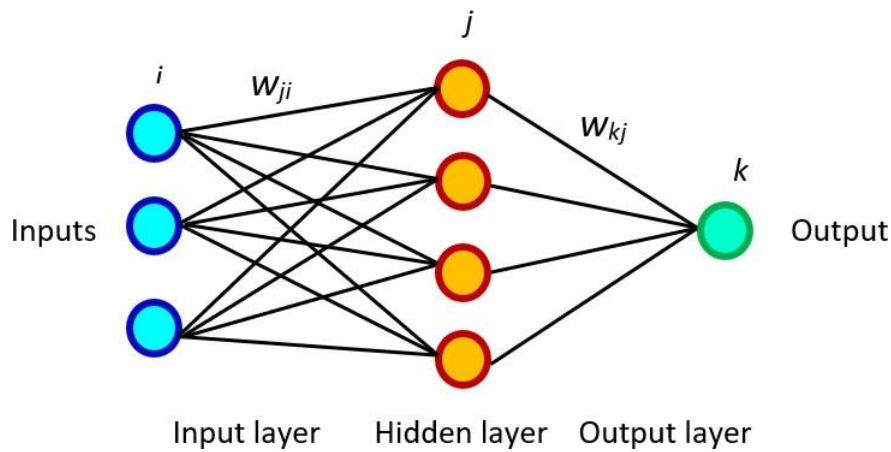


Figure 6. 6. The multilayer perceptron used to differentiate wrist fracture and wrist sprain (Shobayo et al., 2022).

The MLP connections have associated weights that are initially set to random values during training. The weight for the connection to the processing element j from the processing element i is expressed as w_{ji} in Figure 6.6. Similarly, the weight for the connection to the processing element k from the processing element j is expressed as w_{kj} . An MLP learns through multiple iterations using its training file, and then, its performance is evaluated using a test file that contains examples not included in the training file. The operation of an MLP is explained in articles such as (Sharkawy, 2020), but it is very briefly explained here. For each processing element, its associated inputs are multiplied with the corresponding connection weights, and the products are summed. The output of a processing element is obtained by feeding the summed value into a transfer (activation) function. There are several possible transfer functions; however, sigmoid (Harrington, 1993) is commonly used for this purpose. During MLP training, the output of the processing element at the output layer is compared with the provided desired output (fracture = 1, sprain = 0), and the difference is used as the error. The backpropagated learning algorithm (Suliman & Zhang, 2015) is then used to reduce the error iteratively by adjusting the weights associated with the connections.

The MLP used in this study had three inputs (thus three processing elements at its input layer) representing standard deviation from the mean temperature, kurtosis, and interquartile range (IQR) of the 50 selected temperature values (adjusted by subtracting the mean temperature of contralateral ROI). Kurtosis quantifies how heavily the tails of a distribution differ from the tails of a normal distribution. The IQR describes the middle 50% of measures, ordered from lowest to highest. It is the difference between the upper quartile and the lower quartile in an ordered set of measures. The results section provides justification for selecting these three statistical measures.

The MLP had four processing elements for its hidden layer. There is not specific formula to calculate the optimum number of processing elements for this layer, and thus, the decision was based on experimenting with varied number of processing elements. An excessive number of hidden layer processing elements causes poor generalisation (i.e., inadequate performance in correctly identifying participants not included in the training file). Not sufficient processing elements for the hidden layer cause inadequate training. The processing element at the MLP output layer had a range of 0 to 1 (0 representing sprain and 1 as fracture). The receiver operating characteristic (ROC) (Streiner & Cairney, 2007) was used to determine the boundary threshold in this range to differentiate between sprain and fracture. This is further discussed in the results. The MLP training parameters were:

- Error backpropagation learning function to update the weights: gradient descent with momentum. This learning function is commonly used with MLP. The function incorporated two parameters: learning rate and momentum. Learning rate controls its adaptation (learning or training) speed. The momentum term helps the function to move out of local minima to a global minimum when determining error (Qian, 1999). For both parameters, values between 0.01 and 1 were explored, and 0.05 was selected, as it provided the best differentiation.
- Training termination: Training stopped the duration of each trial when the error became insignificant (0.01) or when the number of iterations reached 20,000. The second criteria ensured training to be terminated when the error could not reach its specified target value.
- Transfer (activation function): the sigmoid transfer function was used for all processing elements. It provides an output between 0 and 1 and is commonly used for MLP (Harrington, 1993).

The range (minimum to maximum values) for each measure used as input to the MLP varied. To ensure that the differences in the range of measures did not adversely influence their relative contributions to the MLP output, each measure was individually mapped so that its minimum value corresponded to zero, and its maximum value corresponded to one. Other values were scaled accordingly. The formula used for this purpose was

$$x_s = \frac{x - \textit{Minimum value}}{\textit{Maximum value} - \textit{Minimum value}} \quad 6.6$$

where x_s and x represent scaled values and original values used as inputs to the MLP. This scaling also ensured that the inputs to MLP conformed to the range of the sigmoid transfer function (i.e., 0 to 1).

Given the available sample size, a strategy had to be devised to maximise the scope of the MLP training and accuracy of its performance during evaluation. A limited sample size can often occur in medical machine learning scenarios (Pasini, 2015; Willemink et al., 2020), and thus, strategies to make effective use of the available data were reported (Kokol et al., 2022). In a study, multiple runs of a single instance of a neural network were trained, each run having variations in some training settings, and collective statistics were generated (Shaikhina & Khovanova, 2017). The issue of small sample size to diagnose glaucoma was dealt with by applying a few-shot learning approach (Kim, M. et al., 2017). Oversampling was applied to balance the number of available examples when applying machine learning for chronic kidney disease risk predication (Dritsas & Trigka, 2022). Another approach is data augmentation, where the amount of data is artificially increased by fine adjustments (e.g., flipping, rotation, cropping, zooming, noise addition, colour transformations) of the available data or creating synthetic data (Maharana et al., 2022). The approach has been applied for deep learning of images (Chlap et al., 2021; Li, J. et al., 2019; Shorten & Khoshgoftaar, 2019; Zhang, Chunling et al., 2022). Deep learning has a feature learning capability; however, for the MLP used in this study, a process of feature preparation was undertaken. MLP has a lower data processing capability than deep learning, but it is much less computationally intensive. Therefore, as compared to deep learning, MLP can be applied more quickly and can utilise hardware platforms with lower processing capabilities. The extent an augmentation process may help improving the effectiveness of the MLP in this study (given the complexities of the IR images) is not certain and could be an area of future exploration. This study approached the limited sample size by adapting two investigations.

Investigation A: This involved creation of a training file consisting of 27 randomly selected participants (i.e., 2/3 of the 40 participants) and placing the remaining 13 participants (1/3 of the participants) in a test (evaluation) file. The MLP was trained on the participants in the training file and evaluated on the participants in the test file. The training and test files were again regenerated in a similar manner, and the MLP training and evaluation were repeated. This operation was performed 100 times, and overall, MLP outputs for the test files were averaged. This procedure ensured that there was no bias in selecting specific participants for each of the two files, as the repetitions (trials) resulted in all participants having an opportunity to be included in both the training and tests files.

Investigation B: This was like investigation A, except the number of participants included in the training file was increased to 35, and the number of participants included in the test file

was reduced to 5. The purpose of this investigation was to explore the effect of increasing the number of participants in the training file on the discrimination accuracy of the MLP. In some circumstances, an increase in the training file size may have a positive effect on MLP training, as more participants are examined during each trial. However, its downside is that the MLP evaluation would be on a reduced number of participants.

6.3 Result and discussion

In this section, the features representing the wrist ROI are analysed, and the MLP performance in discriminating between wrist fracture and sprain is presented.

6.3.1 Feature analysis

Average differences between the injured and contralateral uninjured wrists for the statistical measures are provided in Table 6.1. These are also shown as boxplots in Figure 6.7.

Table 6. 1. Average differences between the injured and contralateral uninjured (reference) wrists for the statistical measures, the percentage difference and number of participants differentiated by the statistical measures (fracture (f), sprain (s)).

	Maximum (°C)	Minimu m (°C)	Mean (°C)	Std Dev. (°C)	Median (°C)	Mode (°C)	Skewness	Kurtosis	IQR (°C)
Fracture	1.396	0.696	0.962	0.187	0.938	0.696	0.408	2.478	0.272
Sprain	1.048	0.530	0.711	0.136	0.690	0.530	0.595	2.804	0.202
%Difference	24.942	23.873	26.076	27.322	26.439	23.873	-45.931	13.157	25.752
Number of participants differentiated	13 (f>s)	10(f>s)	12(f>s)	14(f>s)	12(f>s)	8(f>s)	13(f<s)	16(f<s)	14(f>s)

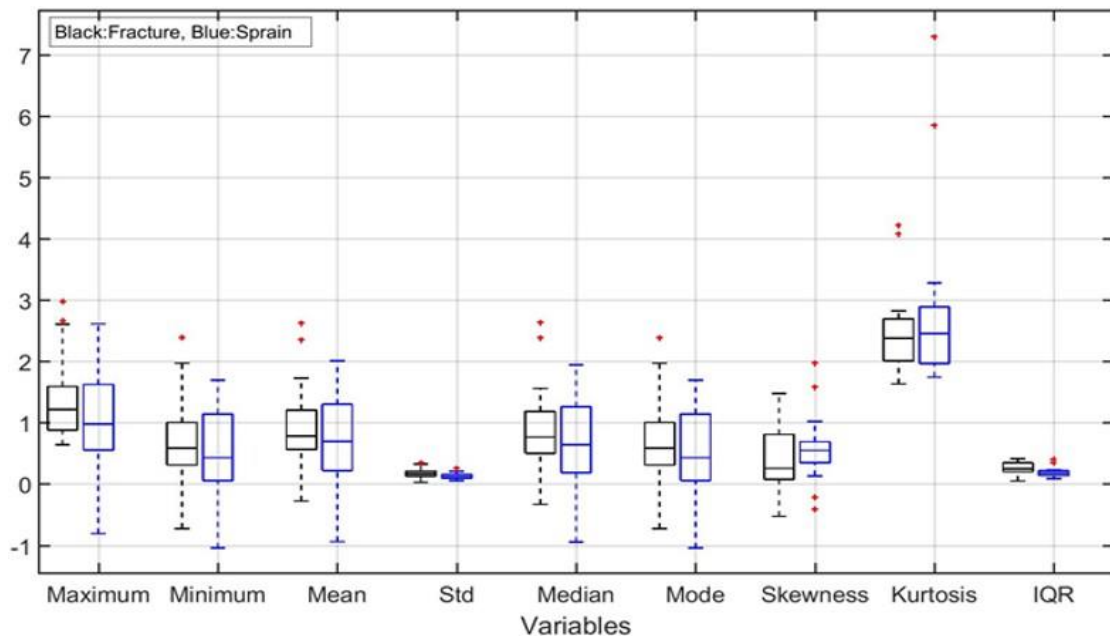


Figure 6. 7. Boxplots of average differences between the injured and uninjured contralateral wrists for the statistical measures. Note: the label for vertical axis is not shown, as the variables do not have the same unit. The unit for minimum, maximum, mean, standard deviation (std) and interquartile range (IQR) is °C. Skewness and kurtosis have no units (Shobayo et al., 2022).

To select the statistical measures that provided greater differentiation between fracture and sprain, the following operations were performed. The mean values of the statistical measures for the sprained wrists were obtained. These values were compared with the values of the associated statistical measure for the fractured wrists. The following observations were made:

- 13 (68.4%) participants with fracture had maximum temperatures greater than sprain participants maximum temperature.
- 10 (52.6%) participants with fracture had minimum temperatures greater than sprain participants minimum temperature.
- 12 (63.2%) participants with fracture had mean temperatures greater than sprain participants mean temperature.
- 14 (73.7%) participants with fracture had standard deviations (from the mean) greater than sprain participants standard deviation.
- 12 (63.2%) participants with fracture had median temperatures greater than sprain participants median temperature.
- 8 (42.1%) participants with fracture had mode temperatures greater than the sprain participants mode temperature.

- 13 (68.4%) participants with fracture had skewness values lower than sprain participants skewness.
- 16 (82.2%) participants with fracture had kurtosis values lower than sprain participants kurtosis.
- 14 (73.7%) participants with fracture had IQR values greater than sprain participants IQR.

The above analysis indicated that the measures' effectiveness to differentiate between the two types of injuries in descending order were:

- Kurtosis
- Standard deviation from the mean and IQR
- Skewness and maximum temperature
- Mean and median
- Minimum
- Mode.

For the measures that were indicative of temperature magnitude (i.e., maximum, minimum, mean, standard deviation, median, mode, IQR), most participants with a fracture had values greater than respective measures for sprain. However, for the measures that were indicative of distribution (i.e., kurtosis and skewness), the opposite was the case.

The above analysis led to the selection of kurtosis, standard deviation from the mean, and IQR for input to the MLP. Inclusion of other measures did not improve the MLP differentiation results, and thus, they were omitted.

6.3.2 Multilayer perception discrimination result for investigation A

A statistical summary of investigation A, indicating the averaged (over 100 trials) MLP outputs for the participants in the test file, is provided in Table 6.2. The averaged MLP outputs for fracture and sprain were 0.589 and 0.349, respectively.

Table 6. 2. Investigation A – averaged multilayer perceptron outputs for participants in the test file (averaged over 100 trials). The values have no units.

Injury Types	Average	Standard Deviation
Fracture	0.589	0.264
Sprain	0.349	0.247

To explore the MLP differentiation effectiveness, the percentage difference (PD) and percentage absolute difference (PAD) were obtained. The formulae used for these were:

$$PD = \frac{F - S}{F} \times 100 \quad (6.7)$$

$$PAD = \left| \frac{F - S}{0.5(F + S)} \right| \times 100 \quad (6.8)$$

Where:

F = MLP output averaged over 100 trials for participants with fracture included in the test file.

S = MLP output averaged over 100 trials for participants with sprain included in the test file.

The PD and PAD values were 40.75% and 45.02%, respectively.

To determine the classification boundary between fracture and sprain, the receiver operating characteristic curve (ROC) was obtained for the averaged MLP outputs (over 100 trials) for the participants in the test file. The resulting plot is shown in Figure 6.8.

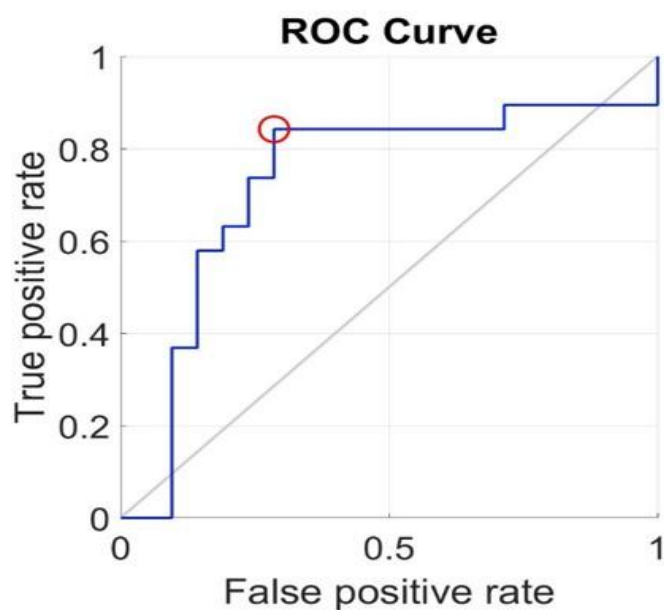


Figure 6. 8. The receiver operating characteristic curve (ROC) for investigation A (Shobayo et al., 2022)

The area under ROC was 0.727, and the ROC optimum point for discrimination for the two types of injuries corresponded to a false positive rate = 0.289 and true positive rate = 0.842 resulting in the threshold value differentiating fracture and sprain to be 0.380. Using this threshold, the averaged MLP output for participants in the test file was analysed, resulting in:

- Number of true positives = 16
- Number of true negatives = 15
- Number of false negatives = 3
- Number of false positives = 6

These measures resulted in discrimination sensitivity of 0.842 (84.2%) and specificity of 0.714 (71.4%). The positive and negative predictive values were 0.723 and 0.833, respectively. The overall discrimination accuracy was 77.5%. The plot of averaged MLP outputs for participants in the test file are shown in Figure 6.9. The threshold level is shown as the horizontal dashed line at 0.380.

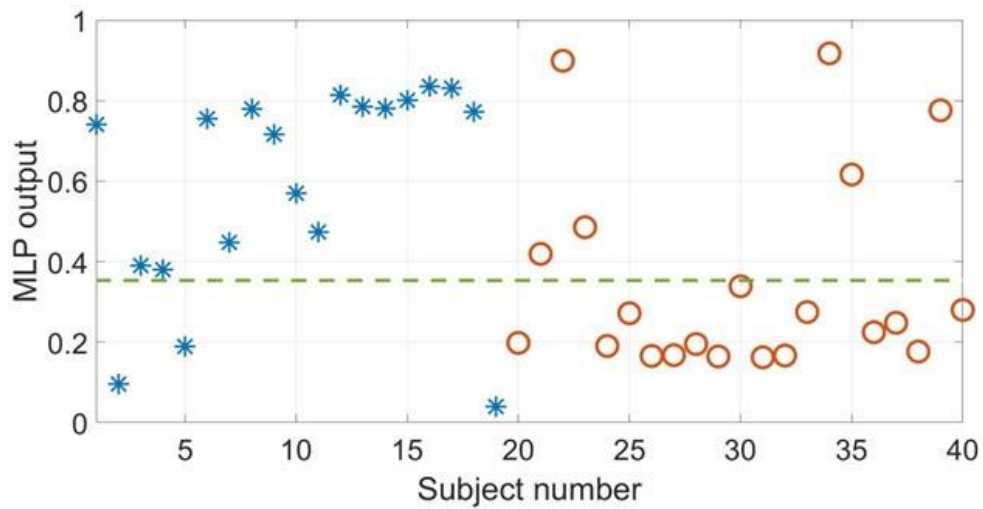


Figure 6. 9. Plot of averaged multilayer perceptron outputs (over 100 trials) for participants in the test file in investigation A. The stars and circles represent participants with fracture and sprain, respectively. The dashed line is the threshold (Shobayo et al., 2022).

Figure 6.9 shows three participants with fracture appearing under the threshold line with sprain participants and six participants with sprain appearing above the threshold line with fracture participants.

6.3.3 Multilayer perceptron discrimination results for investigation B

A statistical summary of investigation B indicating the averaged (over 100 trials) MLP outputs for the participants in the test file is provided in Table 6.3. The average MLP outputs for fracture and sprain were 0.617 and 0.345, respectively.

Table 6. 3. Averaged multilayer perceptron outputs (over 100 trials) for participants in the test file, investigation B. The values have no units.

Injury Types	Average	Standard Deviation
Fracture	0.617	0.280
Sprain	0.345	0.252

The PD and PAD values for this investigation were 44.084% and 56.549%, respectively. To determine the classification boundary between fracture and sprain, the receiver operating characteristic curve (ROC) was obtained for the averaged MLP values (over 100 trials) for the participants in the test file. The resulting plot is shown in Figure 6.10.

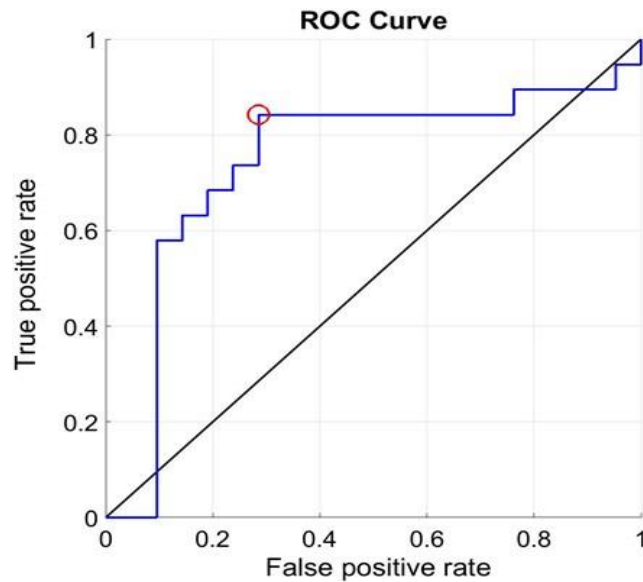


Figure 6. 10. The receiver operating characteristic curve (ROC) for investigation B (Shobayo et al., 2022).

The area under ROC was 0.742, and the ROC optimum point was associated with a false positive rate = 0.286 and true positive rate = 0.842. These corresponded to a differentiation threshold value of 0.353. Using this threshold, the MLP outputs for participants in the test file were analysed, providing:

- Number of true positives = 16
- Number of true negatives = 15
- Number of false negatives = 3
- Number of false positives = 6.

These values resulted in differentiation sensitivity of 0.842 (84.2%) and specificity of 0.714 (71.4%). The positive and negative predictive values were 0.727 and 0.833, respectively. The overall discrimination accuracy was 77.5%. The averaged MLP outputs for participants in the test file are shown in Figure 6.11. The threshold level is shown as the horizontal dashed line at 0.353.

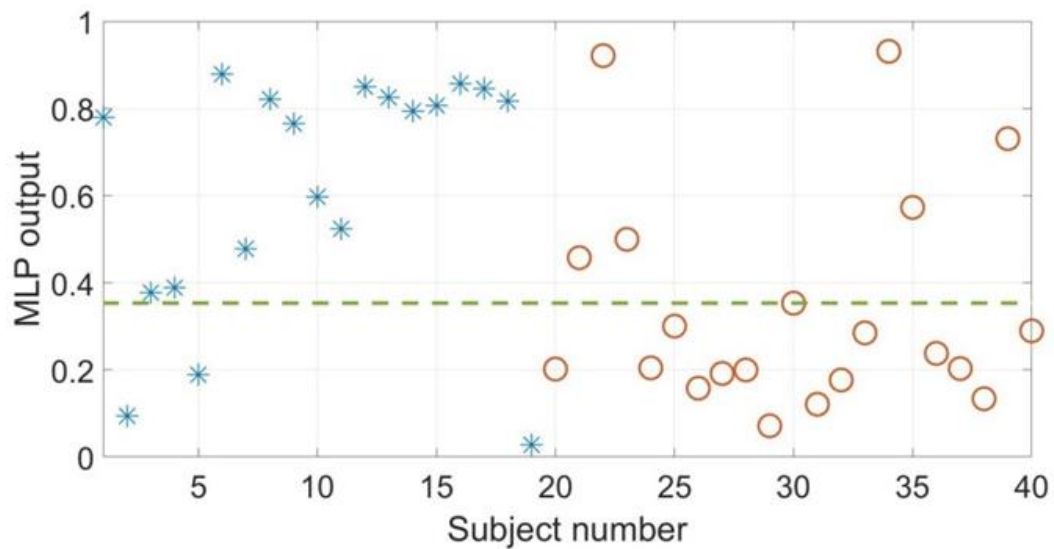


Figure 6. 11. Plot of averaged multilayer perceptron outputs for participants in the test file for investigation B. The stars and circles represent participants with fracture and sprain, respectively. The dashed line is the threshold (Shobayo et al., 2022).

6.4 Summary

The study explored the effectiveness of multilayer perceptron (MLP) and IR thermal imaging (IRTI) to screen for paediatric wrist fractures. A grid structure-based method of representing injured wrists for feature extraction was devised and characterised by statistical measures for input to the MLP. Two investigations were carried out involving different numbers of participants in the training and test files. The sensitivity and specificity obtained by both methods were consistent (84.2% and 71.4%, respectively). The overall accuracy in both methods was 77.5%. The main limitation of the study was the sample size (i.e., 19 with wrist fracture and 21 with wrist sprain). A larger database is likely to improve the differentiation of specificity and specificity. The study indicated that application of MLP to suitably selected IRTI features could have potential for screening for wrist fractures in paediatrics. The method could be explored further for other fractures and in adults.

Chapter 7 – Convolutional Neural Network to Classify Infrared Thermal Images of Fractured Wrists in Paediatrics

7.1 Introduction

In this chapter, a CNN model was devised to identify wrist fractures from IRT images. The ability of the CNN was leveraged to automatically extract features based on the convolution layers as previous methods used manual feature extraction methods. The features extracted by the CNN architecture are automatically learned hierarchical representations of the infrared thermal (IRT) images, rather than manually engineered descriptors. At early convolutional layers, the network captures low-level features such as temperature gradients, edges, and local intensity variations corresponding to thermal contrasts within the wrist region. As the depth of the network increases, it learns mid-level features such as spatial temperature distributions, asymmetry patterns between injured and surrounding tissue, and localised hotspots. In deeper layers, the CNN encodes high-level semantic features, including complex thermal signatures associated with inflammation, vascular response, and injury-specific patterns that distinguish fractures from sprains. A tailor-made CNN algorithm built from scratch, specifically for the IRT images characteristics is presented. An exploration of the effectiveness of image augmentation when developing DL models for IRT image classification was also performed.

Children may sustain wrist fractures from a range of injury mechanisms. They are more at risk of harm from X-ray radiation than adults (Reed et al., 2020). Around 50% of X-ray radiographs obtained to diagnose wrist fractures do not show a fracture. In addition, serial X-rays are often obtained to identify healing fractures where they are not visible initially. Therefore, a harmless, easy-to-use and cost-effective means of identifying wrist fractures at the index presentation could reduce the number of unnecessary X-rays.

Interests in the medical diagnostic and monitoring applications of infrared (IR) thermal (IRT) imaging have grown in recent years, as infrared sensors and image processing techniques have improved significantly. A modern IRT camera produces images by accurately measuring the heat emanating from objects in the form of IR (Guo & Li, 2019). A feature of IRT imaging is its noncontact nature, allowing measurements without disturbance to the individual or risk of contamination (Kastberger & Stachl, 2003). IRT imaging has proved effective in biological applications (Kastberger & Stachl, 2003) and the identification of paediatric bone fractures (Fane De Salis et al., 2018; Reed et al., 2020). An analysis of IRT images demonstrated a statistically

significant temperature difference between the fractured wrists and their contralateral (uninjured) wrists caused by the changes in the blood flow at the injury site (Reed et al., 2020).

Using the ImageNet datasets, which are widely used in industry and academia, CNNs have achieved a remarkable capability for both image classification and pattern recognition tasks. The working principle is usually the same and is composed of convolution layers applied at the input of a supervised DL network that updates its weights of feature maps, learnt from the previous layer, based on their relationships in space to reduce the numbers of parameters needed to train the model (Li, Z. et al., 2021; Liu, Xiangbin et al., 2021; Morid et al., 2021). This is then followed by fully connected layers and the output layer, which could be either classification (SoftMax layer), recognition or clustering.

There are several variants of CNN models commonly referred to as transfer learning (TL) models. These include ImageNet, LeNet, Inception, VGGNet and ResNet (Wu, Meiyin & Chen, 2015). These models have been trained on a very large number of high-resolution images from the internet known as the ImageNet with millions of trainable parameters, with their weights and biases frozen. The input and classification layers can be adjusted to train other types of healthcare-related images such as those from computerised tomography scans, IR thermography and magnetic resonance imaging.

A challenge encountered in applying DL techniques such as CNN to medical applications is small datasets, a problem known as the data challenge (Hernandez-Guedes et al., 2022). A method to deal with it is called data augmentation (Alomar et al., 2023). Data augmentation could allow DL models to improve their data classification performance by artificially generating a greater diversity of training examples (Alomar et al., 2023; Iglesias et al., 2023; Nanni et al., 2021). Therefore, augmentation techniques are transformation operations that modify a sample (e.g., an image) in such a way that it can still be semantically described by the original identity of the sample category or type (Mumuni & Mumuni, 2022). Data augmentation has been applied as part of Raman spectroscopy skin cancer tissue classification (Wu, Man et al., 2021), natural language processing (Li, B. et al., 2022), time series classification with neural networks (Iwana & Uchida, 2021), text classification (Bayer et al., 2022), material microscopic image segmentation (Ma et al., 2023) and image classification (Zhang, Ruirui et al., 2022). There were also studies outlining the effectiveness of generative adversarial networks (Brophy et al., 2023; Goodfellow et al., 2020) in medical image augmentation (Chen, Y. et al., 2022; Frid-Adar et al., 2018; Meor Yahaya & Teo, 2023; Sandfort et al., 2019). A further challenge when using medical data for machine learning diagnostic purposes may be a significant imbalance in the number of

samples (example cases) from different categories. This imbalance may bias the analysis and thus affect the accuracy of the results. A study reported a means to deal with this imbalance by using Euclidean distance between samples of the same category to reduce the associated imbalance (Saatchi & Ramlakhan, 2023).

7.2 Chapter related methodology

The method used in this study comprised data collection and data pre-processing which included fast Fourier transformation (FFT) of the IRT images, image resizing and augmentation. These were followed by classification based on a CNN-based DNN. The operations are outlined in Figure 7.1.

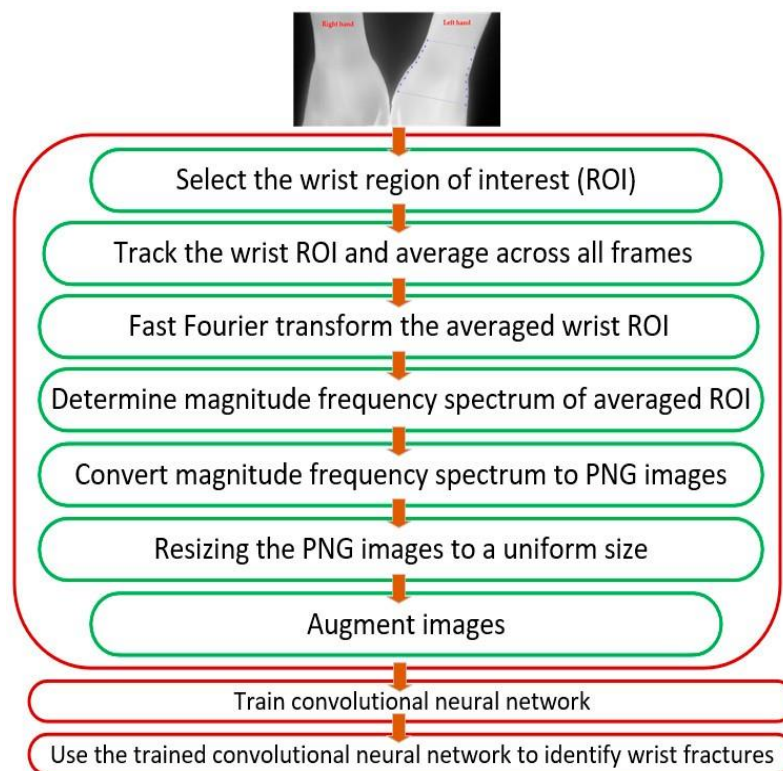


Figure 7. 1. Operations to identify wrist fracture (Shobayo et al., 2024)

7.2.1 Participants' recruitment and data recording

As there can be variations in skin temperature across participants, the mean temperature of the uninjured wrist region of interest (ROI) was subtracted from each pixel value of the injured wrist ROI. The resulting ROIs for the injured wrists were fast Fourier transformed (FFT). Since the wrist ROI varied in dimension across the participants, the FFT process provided a means to resize the images by selecting a section from its centre with a size of 100×100 pixels and thus ensuring they all had identical dimensions. This section had the main frequency components. An IR image participant's hands with a fracture of the left wrist is shown in Figure 7.2a. The fractured left

wrist appears brighter in comparison with the uninjured right wrist, indicating an increase in IR thermal emission. An IR image of a participant's hand with a sprained left wrist is provided in Figure 7.2b. The temperature differences between the injured and uninjured wrist require image processing analyses to interpret their significance.

The ROIs for the injured left wrists are shown by the blue dotted lines on the respective images. A similar ROI was selected from the uninjured wrist. Typical magnitude frequency spectra of the ROI for fractured and sprained wrists are shown in Figures 7.3a and 7.3b, respectively.

The transformed and resized images were then converted into the portable network graphic (PNG) format for CNN input compatibility in preparation for the next stage.

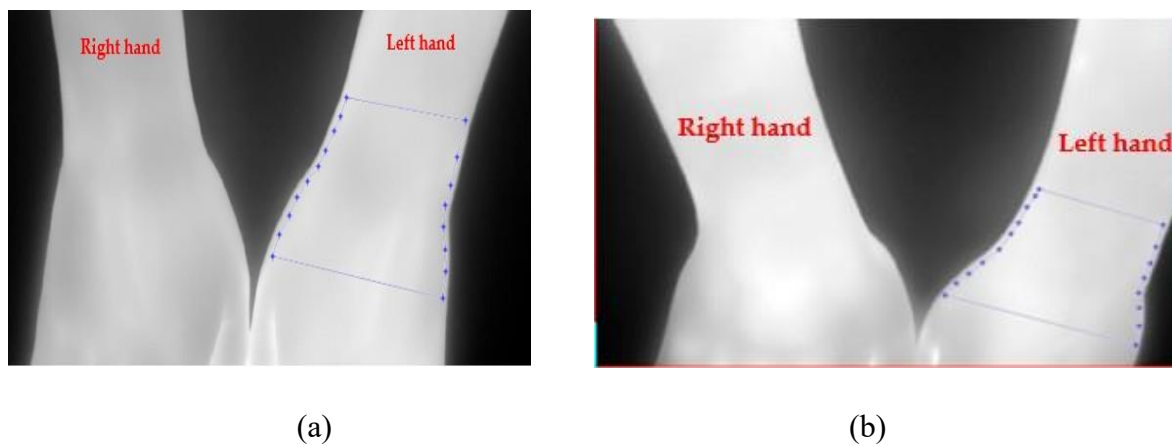
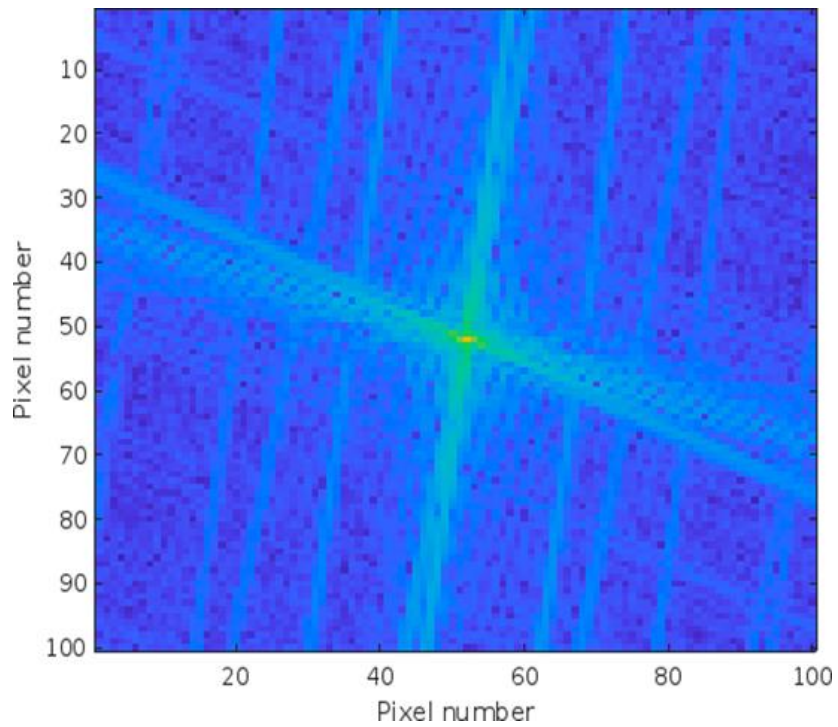


Figure 7. 2. Infrared image of a participants' hands: (a) fracture of left wrist; (b) sprained left wrist. The region of interest is shown by the blue dotted line (Shobayo et al., 2024)

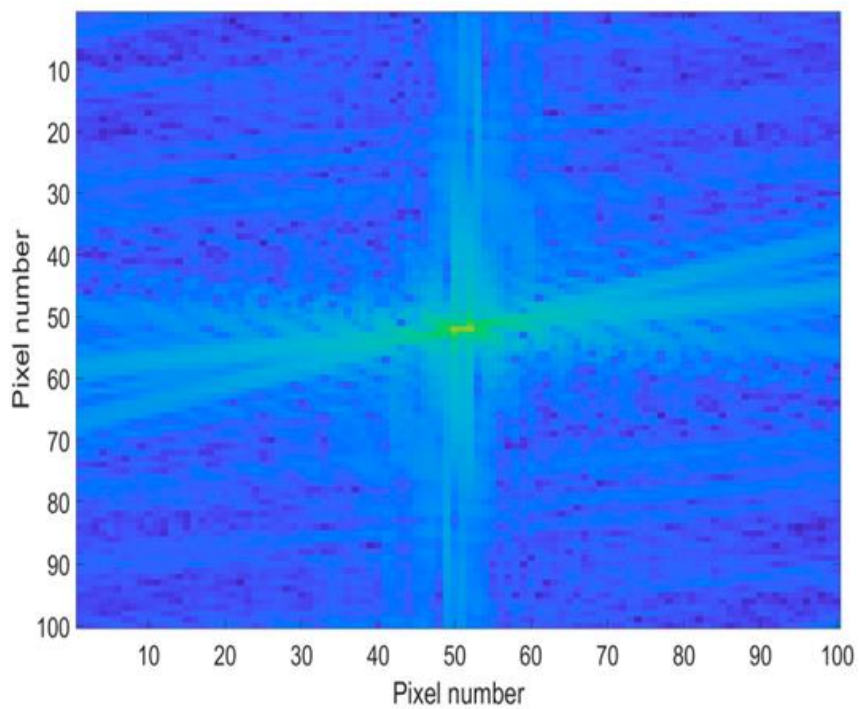
7.2.2 Image augmentation

Geometric transformation methods of rotation, translation, and shearing were performed on the fast-Fourier-transformed and resized ROI sections (images) of the injured wrists. This was to introduce further variabilities in the images to enhance CNN generalisation during training (Shorten & Khoshgoftaar, 2019). Each image was randomly subjected to a single transformation scheme. For those images that were subjected to rotation transformation, the amount of the rotation was randomly selected to be between -90° and 90° . A lower value of rotation was not as effective for classification. For the images that were translated, the amount of translation was randomly selected to be between -3 pixels and $+3$ pixels along the horizontal and vertical axes. The translation allowed the input images to have variations in space and avoid positional bias during training. Finally, for the images that were subjected to shearing transformation, the amount of shearing was randomly selected to be between a shear factor of -2 and 2 , both horizontally and vertically. Shearing affected the shape of the input images by slanting or tilting them in the

specified direction. The values chosen for the translation and shearing were kept quite small as the input data had minimal variations from the point of recording.



(a)



(b)

Figure 7. 3. Magnitude frequency spectra of a typical region of interest for (a) fractured wrist ROI and (b) sprained wrist ROI (Shobayo et al., 2024).

7.2.3 CNN-based deep learning architecture

The computer used to perform the processing operations was an Apple MacBook, with an 8-core CPU, 10-core GPU, 16-core neural engine, 100 GB/s memory bandwidth, and maximum CPU clock rate of 3.49 GHz. The CNN training took 104 min and 41s.

The structure of the CNN-based deep learning neural network architecture is shown in Figure 7.4. The layers with learnable parameters used in the CNN-based DNN architecture for the classification of the images representing fracture and non-fracture are provided in Table 7.1. The hyperparameters used for fine tuning model are provided in Table 7.2.

Table 7. 1. Layer names and their learnable parameters.

Layer Name	Parameters
2D convolutional (Conv_1 (2D))	80
Batch normalisation (Batchnorm_1 (2D))	16
2D Convolutional (Conv_2 (2D))	1168
Batch Normalisation (Batchnorm_2 (2D))	32
2D Convolutional (Conv_3 (2D))	4640
Batch Normalisation (Batchnorm_3 (2D))	64
2D Convolutional (Conv_4 (2D))	18,496
Batch Normalisation (Batchnorm_4 (2D))	128
2D Convolutional (Conv_5 (2D))	73,856
Batch normalisation (Batchnorm_5 (2D))	256
2D Convolutional (Conv_6 (2D))	295,168
Batch Normalisation (Batchnorm_6 (2D))	512
2D Convolutional (Conv_7 (2D))	1,180,160
Batch normalisation (Batchnorm_7 (2D))	1024
2D Convolutional (Conv_8 (2D))	1,180,160
Batch normalisations (Batchnorm_8 (2D))	1024
Fully Connected (FC)	1026

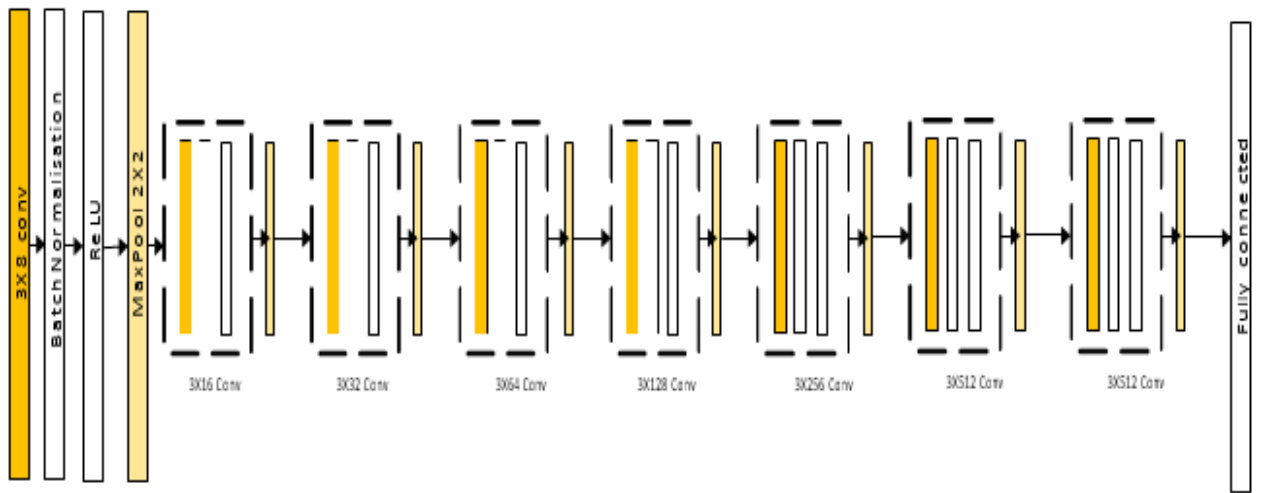


Figure 7. 4 CNN-based deep neural network used in the study (Shobayo et al., 2024).

Table 7. 2. Hyperparameters for CNN based deep neural network.

Hyperparameters	Values
optimiser	Adam
Activation Function	ReLU
Learning rate	0.005
Batch size	170
Max Epoch	250 (no early stopping)
Max Iteration	12250
Loss Function	Cross Entropy

The CNN-based deep neural network consisted of 34 layers in total and 3.9 million learnable parameters, i.e., weights and biases. It comprised a similar combination of convolutional, batch normalisation, rectified linear unit (ReLU) or rectifier activation function (this introduced non-linearity to the deep learning model) and maximum pooling layers, respectively (they helped in the extraction of relevant features from the input images while simultaneously reducing the data dimension).

The CNN-based deep neural network had 8 convolutional layers. The first layer of convolution comprised 8 filters, with a filter size of 3, connected to the input layer. The ‘convolution2dlayer’ from the MATLAB[®] DL toolbox (Version 2023b) was used to create the CNN model. It defined the filter, referred to as the kernel or mask. It used an array of defined weights applied to the neighbouring pixels of the input image while the convolution process was being performed. During the convolution, the filters were arranged at each position of the image input around the

pixels centered in the position. There was an element-wise multiplication of the neighbouring pixels of the input image, and the results of the multiplication were then summed to generate the filter output. This was then slid through the entirety of the image with the convolution process at each position of the window. The output image, also known as the feature map, had the same dimension as the input image. Its values were determined by the convolution operation.

As the layers deepened, the number of convolutional filters increased by $2(n+2)$, where n is the CNN convolution layer number, until it reached 512 filters, which was the maximum number of filters that could be achieved based on the size of the input image, as the choice of the convolutional filters must be less than the input image size. As the layer number increased, it enhanced the ability of the CNN model to distinguish features of the images (Venkatesan & Li, 2017).

A batch normalisation layer followed each convolutional layer. The batch normalisation layer sped up the CNN convergence by normalizing the values of the calculated weights and biases from the previous convolution layer to have a mean value close to zero and a standard deviation close to 1 before passing to the next layer of the CNN.

A rectified linear unit (ReLU) activation layer followed each batch normalisation layer. The choice of the ReLU activation function was based on the non-linear classification task that was required of the input images. Each ReLU layer ensured that neurons with negative values remained inactive, allowing neurons with positive values to be activated for the next convolution layer. This helped the CNN model to learn only the main image features, thereby reducing overfitting by the model during its training (Nair & Hinton, 2010).

The final convolution layer was connected to a fully connected (FC) layer of processing neurons with two outputs representing the injury types, i.e., fracture and non-fracture. The output vectors from the FC layer comprised positive and negative values.

The next layer was the SoftMax layer that converted the output from the FC layer into normalised class probabilities (Zhu, Qiuyu et al., 2020). The classification layer generated the classification of each image based on the probability of injury being a fracture or a non-fracture (sprain) from the SoftMax layer.

The CNN model was trained on 250 epochs (each epoch was 49 iterations) without early stopping, where the final model metrics were calculated based on the average validation metrics throughout the specified number of epochs. The training process was examined for early stopping at 60 epochs (corresponding to 2940 iterations) where there was a global minimum in terms of

the training error. As, at that point, the differentiation accuracy was not sufficiently high, the training was allowed to continue for 250 epochs. Thereafter, there was no further improvement in the CNN classification accuracy.

Figures 7.5 and 7.6, respectively, show the training and loss (i.e., error) performances of the CNN model with the validation dataset. The blue graph in Figure 7.5 represents the training of the model, with the black graph representing the model's validations using the validation data with validation performed after 250 epochs. The red graph in Figure 7.6 shows the training loss and the black graph shows the validation error.

7.2.4 Evaluation metrics

The efficacy of the proposed CNN-based DL neural network was assessed using the confusion matrix, which provided classification accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). These metrics were obtained from the model's classification of true negatives (TNs), true positives (TPs), false positives (FPs) and false negatives (FNs). The receiver operating characteristic (ROC) was also used to evaluate the model, providing the area under curve (AUC) for clinical use.

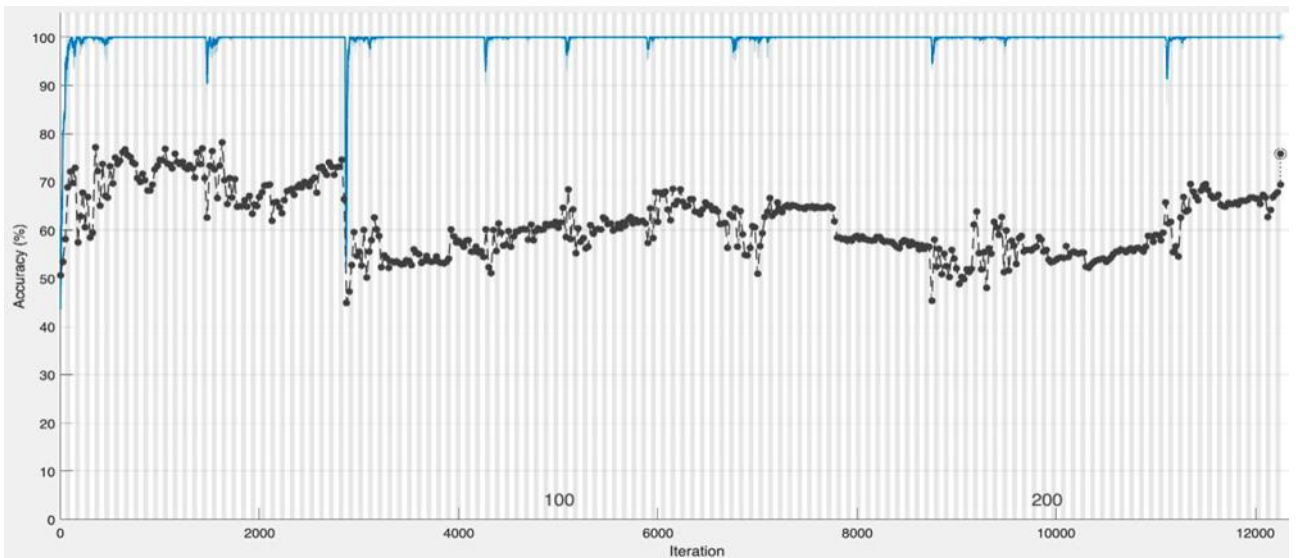


Figure 7. 5. The CNN model training with validation data. The blue graph represents the model's training, and the black graph represents the validation using the validation dataset with validation carried out after 250 epochs (Shobayo et al., 2024).

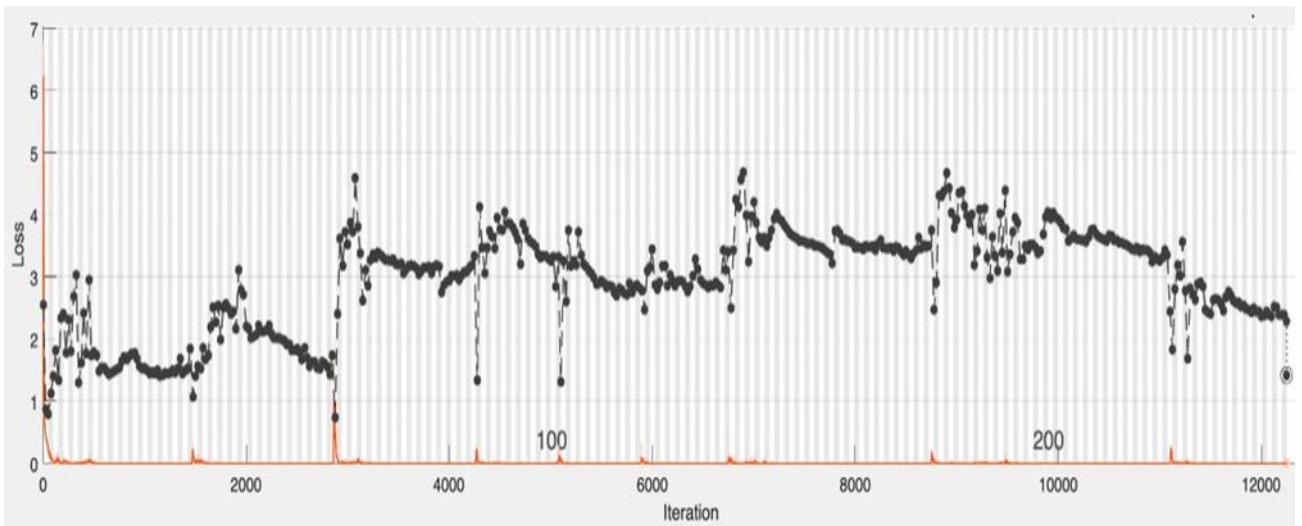


Figure 7. 6. The CNN model loss with validation data. The red graph represents the model training loss (or error), and the black graph represents the validation error using the validation dataset with validation carried out after 250 epochs (Shobayo et al., 2024).

7.3 Results

Two variants of the CNN-based DNN architecture were investigated. The first variant performed classification with image augmentation without a dropout layer, and the second variant did not include augmentation but had a dropout layer.

When training a DL model on small datasets, there is a tendency for overfitting to occur due to the large number of layers used in the network. To deal with this effect, techniques such as augmentation (Pham et al., 2018) and regularisations (dropout) (Wu, Haibing & Gu, 2015) have been reported. In the first experiment, the images were randomly rotated, translated, and sheared as described in section 7.2. For the second experiment, a dropout layer was added to the to the last convolution layer, just after the maximum (max) pooling layer and before the fully connected (FC) layer. A value of 0.2 was selected for the dropout layer, meaning that 20% of the nodes were dropped with every update in weight during training. A larger value of dropout would have negatively affected the performance of the model as more nodes would have been dropped.

Binary classifications were provided for fractures and non-fractures (sprains) and presented for both experiments. The dataset comprised 11,960 (i.e., 40 participants \times 299 images) fast-Fourier-transformed IRT images of 40 participants, 21 without a fracture (sprain) and 19 with a fracture. The dataset was split into 70% of the images for training (13 participants with a fracture and 15 participants without a fracture, totalling 8372 images) and 30% of the images for validation (6

fractures and 6 non-fractured, totalling 3588 images). The number of images for training and validation is summarised in Table 7.3.

Table 7. 3. Number of images used for CNN training and validation.

Injury	Training	Validation
Fracture	3887	1794
Sprain	4485	1794

7.3.1 CNN-based deep neural network with augmentation and without dropout

The CNN-based deep neural network model was configured based on the hyperparameters provided in Table 7.2. This model was constructed using the MATLAB[®] DL toolbox. To indicate the effectiveness of the model, the average values of sensitivity, specificity, NPV, PPV and accuracy are provided in Table 7.4, and the confusion matrix is provided in Figure 7.7. The receiver operating characteristic curve (ROC) (Obuchowski & Bullen, 2018) is shown in Figure 7.8 with the model operating points for fracture and non-fracture. The CNN-based deep neural network model achieved an area under the ROC curve (AUC) of 0.82.

Table 7. 4. Evaluation metrics for CNN-based deep neural network with augmentation and without dropout.

Injury Type	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)	Accuracy (%)
Fracture	88.3	68.3	80.4	72.5	75.8

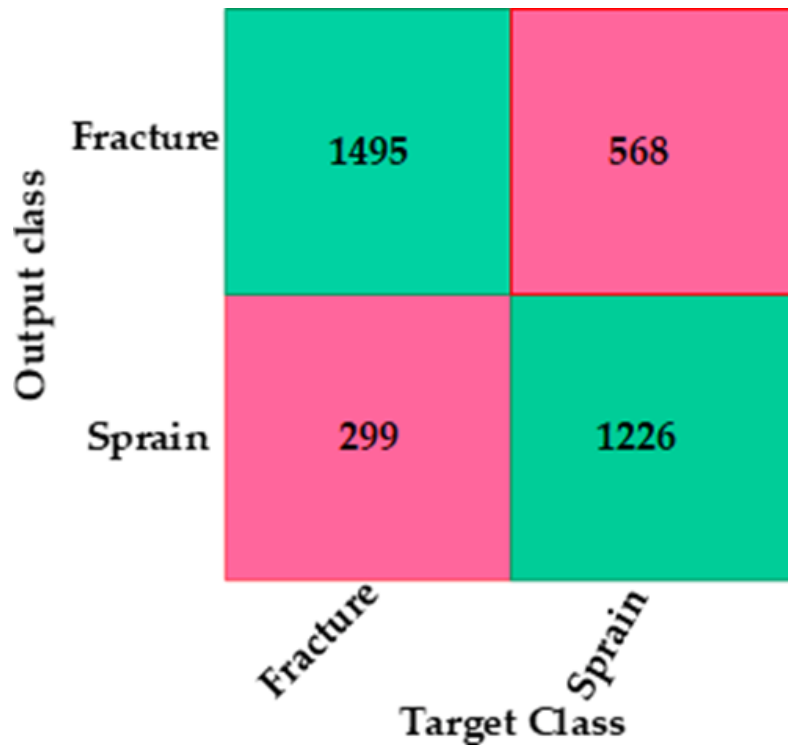


Figure 7. 7. Confusion matrix for CNN-based deep neural network for wrist fracture versus sprain. The squares contain the number of images for validation dataset, total: 3588 images (Shobayo et al., 2024).

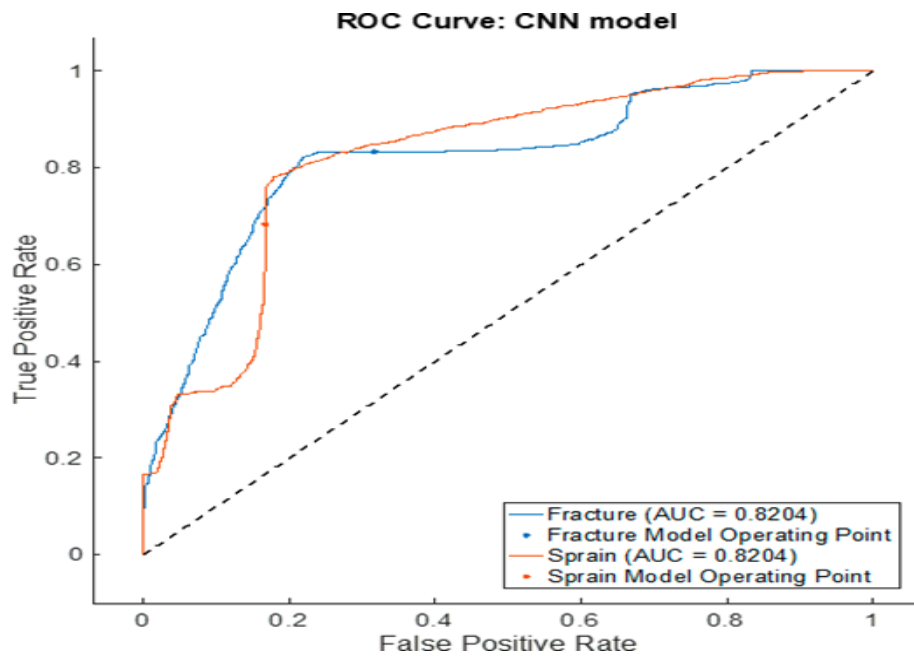


Figure 7. 8. Receiver operating characteristic curve for CNN-based deep neural network (Shobayo et al., 2024).

7.3.2 CNN-based deep neural network without augmentation and with dropout

In this part of the study, the same hyperparameters and CNN architecture was used, but without image augmentation. To compensate for overfitting, a dropout layer (0.2) was introduced, as discussed earlier. The sensitivity, specificity, NPV, PPV and accuracy are provided in Table 7.5.

Table 7. 5. Evaluation metrics for CNN-based deep neural network without augmentation and without dropout.

Injury	Sensitivity (%)	Specificity (%)	NPV (%)	PPV (%)	Accuracy (%)
Fracture versus sprain	50	66.7	60.0	57.1	58.3

7.4 Summary

This study developed a CNN-based deep neural network architecture to differentiate fractured and non-fractured wrists using infrared thermal (IRT) images of 40 participants (11,960 images). The IRT images were fast Fourier transformed to aid their representation and resized in preparation for the CNN model. Augmentation was performed to improve the generalisation of the model during training, thereby reducing overfitting by the model. The model performance was examined in two scenarios. With image augmentation and no dropout, the model achieved accuracy, sensitivity, specificity, NPV and PPV values of 76%, 88%, 68%, 80% and 73%, respectively. An ROC-AUC of 0.82 was also obtained. For the second scenario, where regularisation of the model was performed by including a dropout (0.2) layer for the CNN-based deep neural network architecture, the values of the evaluation metrics decreased. This study was successful in using artificial intelligence-assisted IRT images to identify wrist fractures.

Chapter 8 – Adaptive Neuro-Fuzzy Inference System (ANFIS) Framework for Paediatric Wrist Injury Classification

8.1 Introduction

Adaptive Neuro-Fuzzy Inference Systems (ANFIS) represent a powerful fusion of two prominent artificial intelligence paradigms: fuzzy logic and artificial neural networks (ANNs). It is a supervised learning methodology that integrates the adaptive learning strengths of artificial neural networks with the human-interpretable rule-based reasoning of fuzzy logic (Suparta & Samah, 2020). By marrying the capacity of ANNs to learn complex nonlinear mappings from data with the interpretable, rule-based reasoning of fuzzy inference, ANFIS delivers both predictive power and transparency. Like a conventional neural network, ANFIS is trained using sample datasets: during the “learning” phase, network parameters are adjusted to minimise prediction error on a training set. Once trained, the resulting model structure which comprises both antecedent (fuzzy membership) parameters and consequent (linear output) coefficients is evaluated on previously unseen test data (Talpur et al., 2022). Low errors on this validation set confirm that the ANFIS architecture is well-suited to the problem at hand. A notable criticism of standard neural networks is their “black-box” nature: the learned connection weights do not readily translate into human-understandable rules. ANFIS overcomes this limitation by embedding a fuzzy inference system within the network framework. Each neuron corresponds to a fuzzy rule (“if – then” statement), and its parameters can be interpreted as linguistic thresholds. As a result, once training is complete, the entire model can be expressed as a compact set of fuzzy rules which will help facilitate knowledge extraction, validation, and expert refinement (Kar et al., 2014).

Technically, ANFIS implements a Sugeno-type fuzzy inference model via an adaptive network of nodes and weighted links. Training proceeds in two intertwined phases employing a hybrid algorithm: in the forward pass, given fixed premise (membership) parameters, consequent coefficients are computed by least-squares estimation to best fit the data (Hamdan & Garibaldi, 2010). In the backward pass, the resulting error gradients propagate back through the network, and the premise parameters (defining each Gaussian or bell-shaped membership function) are tuned via gradient descent (Roy, 2005; Saatchi, 2024). By decomposing the optimisation into these two complementary steps, ANFIS reduces the dimensionality of the search space compared to pure back-propagation, yielding far faster convergence and more stable parameter estimates.

8.1.1 ANFIS Model development for fracture prediction.

The ANFIS framework has been successfully applied to distinguish between fractures and non-fracture injuries (such as sprains) from the IRT images of the subjects. One of ANFIS's primary benefits lies in its hybrid architecture: the neural network component automatically adjusts weights and biases during training to minimise prediction error, while the fuzzy logic component establishes a set of membership functions and corresponding fuzzy rules that capture expert knowledge about how temperature patterns relate to injury status. This dual mechanism allows the model to learn complex, nonlinear relationships from data while retaining a transparent rule structure. As a result, ANFIS can achieve more efficient training, better generalisation to new cases, and enhanced "memory" of patterns it has encountered. By marrying gradient-based optimisation with linguistic, rule-based inference, ANFIS delivers robust performance on challenging diagnostic tasks such as thermal image-based fracture detection.

8.2 Chapter related methodology

In this section, the ANFIS methodology used to distinguish between fracture and sprain of the IRT images is presented.

8.2.1 ANFIS model development

Most ANFIS implementations, regardless of their specific application, share a common five-layer architecture (Jang, 1993). In this study, we adopted this standard five-layer configuration (see Figure 8.1) to leverage its ability to blend human-readable fuzzy rules with adaptive neural learning. Each of the five layers serves a distinct purpose and is governed by its own mathematical transform.

Layer 1: Fuzzification (Premise Parameters) layer

The first layer accepts raw inputs, also referred to as crisp input which is the infrared thermal pixel values and converts each crisp value into one or more fuzzy membership grades between 0 and 1. This process, known as fuzzification, is achieved by passing each input x_j through a set of membership functions (MFs). In our case we used Gaussian MFs, each parameterised by a centre c and width (δ). These centres and widths are sometimes referred to as the *antecedent* or *premise* parameters, since they define the "if" side of each fuzzy rule ("If temperature is high, then...").

Layer 2: Rule firing strength (Product Layer)

In the second layer, each neuron corresponds to one fuzzy rule and receives the fuzzified grades from Layer 1. It computes the rule's *firing strength* w_j by taking the product of its incoming

membership grades. Mathematically, for rule i with antecedent or premise MFs $\mu_{i1}(x_1), \mu_{i2}(x_2), \dots, \mu_{in}(x_n)$ the output is

$$w_i = \prod_j \mu_{ij}(x_j) \quad 8.1$$

Layer 3: Normalisation

The third layer normalises these raw firing strengths so that they sum to one. That is, each normalised weight is given by:

$$\bar{w}_i = \frac{w_i}{\sum_k w_k} \quad 8.2$$

By normalising, we ensure that the subsequent weighted averaging in Layer 5 remains bounded and comparable across different input scenarios.

Layer 4: Rule consequent (defuzzification) layer

This layer implements the *consequent* side of each fuzzy rule using a TSK (Takagi – Sugeno – Kang) formulation. Each rule has its own linear function

$$f_i^1(x) = \sum_j p_{ij} x_j + p_{i0}, \quad f_i^2(x) = \sum_j q_{ij} x_j + q_{i0} \quad 8.3$$

where the coefficients p_{ij} and bias p_{i0} , q_{ij} and bias q_{i0} are the *consequent parameters*, learned during training. Layer 4 multiplies each rule's normalised firing strength \bar{w}_i by its consequent output $f_i(x)$.

Layer 5: Output Aggregation

Finally, the fifth layer aggregates all rule contributions by summing them:

$$o_1 = \sum_i \bar{w}_i f_i^1(x), \quad o_2 = \sum_i \bar{w}_i f_i^2(x) \quad 8.4$$

This output o_1, o_2 is the model's crisp decision for where in our application, a continuous score indicating fracture versus sprain likelihood. For this study the outputs are a single continuous values/scores indicating the likelihood of fracture versus sprain. We set a threshold of 0.5 to help differentiate between fracture and sprain.

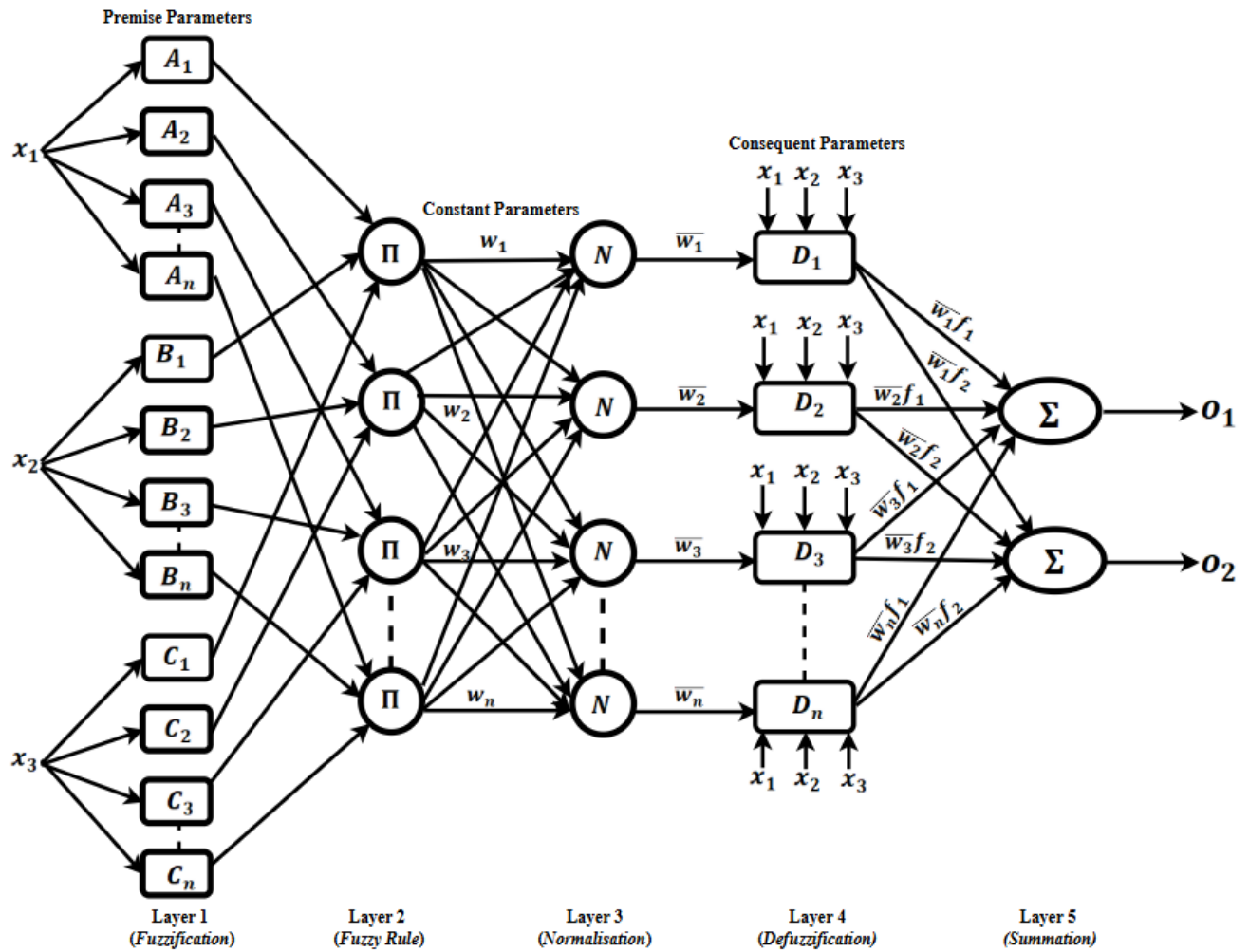


Figure 8.1. 3-input 2-output ANFIS architecture of the fracture classification system (Shobayo et al., 2025).

The ANFIS network in this study was configured to learn from three carefully chosen thermal image descriptors: the standard deviation (SD), the interquartile range (IQR), and the kurtosis of the temperature distribution across each wrist region. These three features were selected specifically because, when the temperature difference between an injured wrist and its healthy counterpart for every subject was measured, SD, IQR, and Kurtosis consistently exhibited the largest separations. In other words, these statistics differed most markedly between injured and uninjured areas, making them the strongest predictors of fracture versus sprain (no-fracture) in our infrared thermography data.

By contrast, six additional candidate descriptors were examined which were maximum temperature (Max), minimum temperature (Min), mean temperature (Mean), median temperature (Median), most frequent temperature (Mode), and skewness of the temperature distribution, but found that none of these variables delivered reliable discrimination. Their values overlapped

heavily between injured and uninjured hands, failing to offer the clear signal needed for accurate classification by the ANFIS model. Therefore, to maximise predictive accuracy and reduce model complexity, the ANFIS input layer were restricted to only SD, IQR, and Kurtosis. Details of this selection is discussed in chapter 4.

8.2.2 Cross validation

A five-fold cross-validation procedure ($K=5$) was employed to systematically tune two critical hyperparameters for the ANFIS model: the Gaussian membership-function width (σ) and the total number of fuzzy rules. In each of the five splits, the training subset was used to fit the model under a candidate combination of σ and rule count, and the remaining validation fold measured performance. By averaging the validation errors across all folds, the sigma (σ) value and rule-set size that yielded the lowest mean squared error and best generalisation across the entire dataset was identified. For all experiments, the Gaussian membership-function form was selected. This bell-shaped MF is widely used in the fuzzy-systems literature because it provides a smooth, continuous mapping from crisp inputs into the $[0,1]$ fuzzy degree, and it is defined by just two parameters: a central location c and the width σ . This simplicity facilitates both transparent rule interpretation and efficient parameter learning. Although Gaussian MFs were the primary choice, a variety of alternative shapes are available and sometimes preferable depending on the application. These include trapezoidal and triangular functions, which offer piecewise-linear transitions. Others include the generalised bell, sigmoid, and other parametric curves (Saatchi, 2024).

The ANFIS model was developed with Python and was trained for 4,000 epochs, but an early-stopping criterion to guard against overfitting was also implemented. Once the mean squared error (MSE) on the training set dropped below 0.08, training should be stopped immediately. This strategy ensures that the model does not continue learning noise once an acceptable error threshold has been reached. The IRT images were trained on the first-order Sugeno–Kang ANFIS architecture, which can be viewed as a feedforward neural network whose weights and biases are optimised via gradient-descent. In this context, each fuzzy rule corresponds to a set of membership-function parameters in the antecedent layer and a linear function in the consequent layer, all of which are tuned simultaneously during training.

To find the best configuration, grid search was performed over two key hyperparameters. First, we experimented with four different fuzzy-rule counts – 2, 3, 6, and 9 rules to determine how many rules provided the best trade-off between model complexity and predictive accuracy. Secondly, three different values for the Gaussian membership-function width parameter, σ : 0.1,

0.2, and 0.3 were also used for testing. Each combination of rule count and σ was evaluated using five-fold cross-validation, with the configuration yielding the lowest average MSE on held-out folds selected as optimal. The combination that produced the minimal validation error consisted of nine (9) fuzzy rules paired with a Gaussian MF width of $\sigma = 0.1$.

8.3 Results and discussion

All input features i.e., standard deviation (Std), interquartile range (IQR), and kurtosis were first normalised to a [0, 1] interval so that each variable contributes equally during training and avoids dominance by any single metric. After normalisation, the full dataset was split into two subsets: 80 percent of the samples were used to fit the ANFIS model, while the remaining 20 percent were held out for independent testing and performance evaluation. Training proceeded using a plain-vanilla gradient-descent algorithm: we set the learning rate to 0.05 and allowed up to 4,000 epochs of iterative weight and parameter updates. Prior to this, a grid search had identified the best combination of Gaussian-membership-function width (σ) and fuzzy-rule count based on cross-validation error. Once those optimal hyperparameters were determined, they were fixed and used in the final ANFIS training run. This process ensured that the model not only learned effectively from the normalised three-feature inputs but also leveraged the most appropriate membership-function geometry and rule complexity for accurate fracture versus sprain classification.

In an ANFIS model, the process of inferencing can be understood as a “divide and conquer” strategy applied to the n-dimensional input space. Each fuzzy rule’s antecedent (“IF” clause) effectively partitions that continuous input space into smaller, overlapping regions or clusters. When a new input arrives, its degrees of membership in each of these regions are calculated, and the rule’s consequent (“THEN” clause) computes a local output. Finally, all local outputs are combined, typically via a weighted average based on rule firing strengths to produce a single, global prediction. Thus, the structure and performance of the entire ANFIS depend critically on how the input space has been divided into these fuzzy clusters and how accurately those clusters capture the underlying data distribution.

To optimise the placement of these fuzzy-set centres, a “scatter partitioning” approach was adopted, which was originally proposed by (Yeom & Kwak, 2018). Rather than simply seeding membership functions at random locations, this method uses a clustering algorithm to discover natural groupings in the training data. In our implementation, each input dimension was partitioned into n clusters, and each cluster centre became the parameter for a corresponding membership function. The total number of fuzzy rules then becomes exactly the number of

clusters raised to the power of the number of inputs – a direct consequence of taking every possible combination of cluster indices across dimensions.

Experiment was carried out on two distinct clustering techniques for identifying those cluster centres: traditional K-means and fuzzy C-means. K-means yields hard, non-overlapping clusters by assigning each data point to its nearest centre, whereas fuzzy C-means produces a soft partition, allowing each point to belong to multiple clusters with varying degrees of membership. In addition, we compared these data-driven initialisations against a baseline in which Gaussian membership functions were centred at random values. By running three separate experiments i.e., random Gaussian seeds, K-means-derived centres, and fuzzy C-means-derived centres, we were able to evaluate the impact of each initialization strategy on both the speed of convergence during training and the ultimate predictive accuracy on held-out data. In summary, the pipeline of the developed ANFIS model is presented in the algorithm shown next page.

Algorithm: ANFIS model selection and evaluation

Inputs: dataset $D = (X, y)$; hyperparameter grids $\Sigma = \{0.1, 0.2, 0.3\}$ (Gaussian membership width), $R = \{2, 3, 6, 9\}$ (rule count); early stopping rule: stop if train MSE < 0.08 or max epochs.

Outputs: Selected (σ^*, r^*) ; final ANFIS; one-time test metrics.

Step 1: Outer hold-out split (once).

Stratify and split $D \rightarrow D_{train}(80\%), D_{test}(20\%)$. Lock D_{test} until Step4.

Step 2: Inner 5-fold cross-validation on D_{train} .

For each $(\sigma, r) \in \Sigma \times R$:

Partition D_{train} into 5 stratified folds.

For each fold k :

Fit scaler on fold-train; transform fold-train/val.

Initialise MF centres on fold-train only (using r)

Train ANFIS with width σ , early stopping (MSE < 0.08).

Score on fold-val (val MSE)

Average fold scores $\rightarrow \text{score}(\sigma, r)$.

Select $(\sigma^*, r^*) = \arg \min_{(\sigma, r)} \text{score}(\sigma, r)$

Step 3: Final Fit on full 80% train.

Fit scaler on all D_{train} ; transform train and later (test).

Initialise MF centres on all D_{train} via K-means /FCM with r^* .

Train ANFIS with σ^* , early stopping.

Step 4: Single evaluation on test set.

Transform D_{test} with the train-fitted scaler; predict once; compute final metrics.

8.3.1 Experiment 1: Gaussian membership function based on random normal centre values.

In our initial experiment, the classic ANFIS approach was adopted by initialising each membership-function centre c with random values drawn from a Gaussian distribution. During training, these centres, along with the network's consequent weights, are refined through gradient-descent back-propagation to minimise prediction error. At the core of this method is the Gaussian membership function, which translates each crisp input into a degree of membership via a smooth, bell-shaped curve. The curve's location along the input axis is determined by the centre parameter c , effectively acting as the mean of the Gaussian. Its spread, or standard deviation, controls how quickly membership falls off away from the centre. Based on prior cross-validation experiments, we fixed this standard deviation to $\sigma = 0.1$. By combining randomly initialised centres with a carefully chosen width, we ensure that the network begins with diverse, well-distributed fuzzy sets and then iteratively adapts both the premise (centres) and consequent parameters to capture complex, nonlinear relationships in the data.

Table 8. 1. Summary of performance for training and validation targets with epochs(t) for experiment 1

Performance Measure	Generalised Membership Function (Generalised bell with random centre)					
	Epoch 800	= Epoch 1600	= Epoch 2400	= Epoch 3200	= Epoch 4000	=
MSE Training	0.1534	0.1407	0.1352	0.1323	0.1304	
MSE Validation	0.1925	0.1729	0.1611	0.1544	0.1514	
RMSE Training	0.3917	0.3752	0.3677	0.3637	0.3612	
RMSE Validation	0.4387	0.4158	0.4014	0.3929	0.3891	
Validation -Training Δ MSE	0.0390	0.0321	0.0259	0.0221	0.0209	
Elapsed Time (seconds)	38.9	77.7	116.4	155.2	194.1	

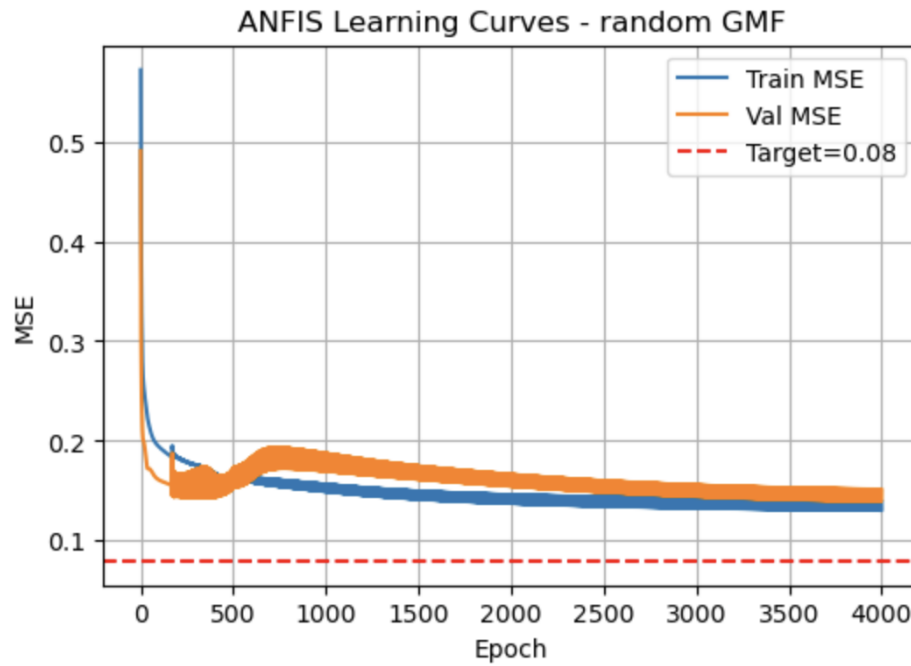


Figure 8. 2. Training and Validation MSE for Experiment 1 (Shobayo et al., 2025)

The plot in Figure 8.2 shows how both training (blue) and validation (orange) mean squared error (MSE) steadily decline over the 4,000 epochs when using randomly initialised Gaussian membership functions. Initially the MSE drops quickly, then tapers off, approaching but never quite reaching the red target line at 0.08 for training MSE. The validation curve dips below the training curve around epoch 300, then rises slightly before converging back toward the training error, which indicates some over- and under-shooting as the model settles.

The Table 8.1 quantifies performance every 800 epochs. By epoch 800, training MSE is 0.1534 and validation MSE 0.1925 (RMSEs 0.392 and 0.439), with a 0.0390 gap. By epoch 4,000, errors have further shrunk (training MSE = 0.1304, validation MSE = 0.1514) and the MSE gap narrows to 0.0209. RMSE follows the same downward trend. Elapsed time grows linearly at about 39 s per 800 epochs, totalling roughly 194 s at 4,000 epochs.

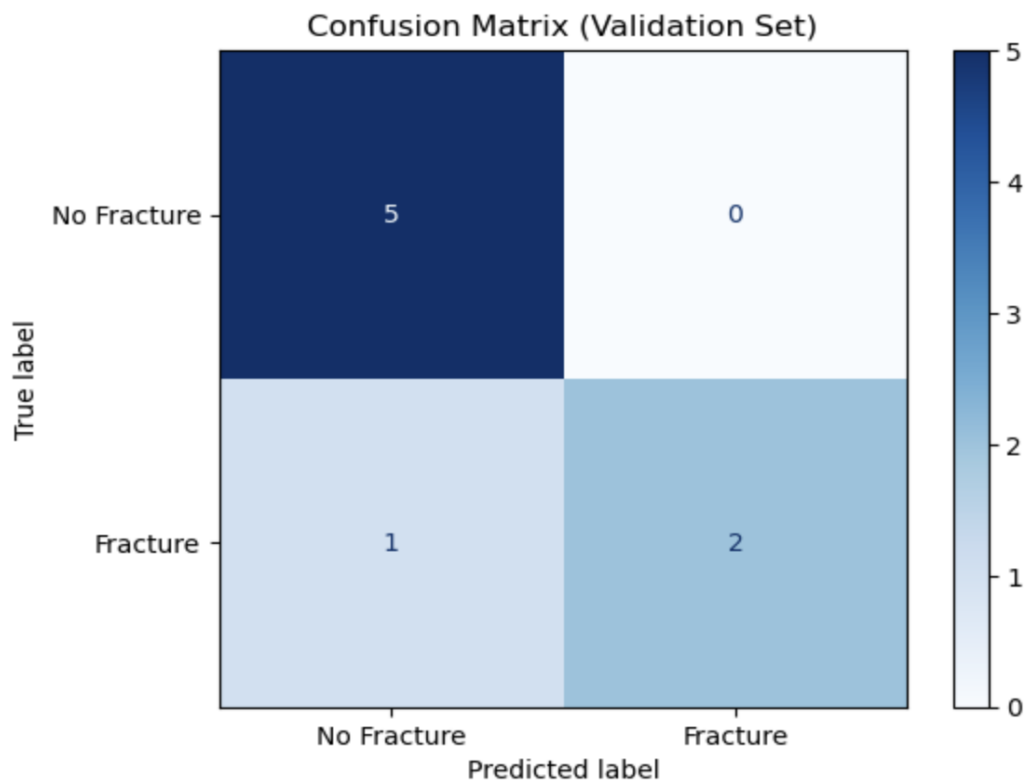


Figure 8. 3. Confusion Matrix of Experiment 1(Shobayo et al., 2025)

The confusion matrix in Figure 8.3 summarises the ANFIS model’s performance on the validation set. The rows represent the true labels (“No Fracture” and “Fracture”), while the columns show the model’s predicted labels.

- True Negatives (No Fracture correctly predicted): 5
- False Positives: 0
- False Negatives (Missed fractures): 1
- True Positives (Fractures correctly predicted): 2

Out of eight cases, the model correctly identified seven (5 non-fractures and 2 fractures) and missed one fracture. This indicates high specificity (no healthy wrists misclassified) and a small false-negative rate, suggesting the model reliably flags non-fractures but may occasionally overlook actual fractures.

The AUC-ROC is shown in Figure 8.4, evaluates the ANFIS model’s ability to distinguish fractures from non-fractures on the validation set. The true positive rate (sensitivity) is plotted against the false positive rate at various classification thresholds. The orange ROC line hugs the top and left axes, indicating that the model achieves 100% sensitivity with 0% false positives

across thresholds which is set at 0.5. The dashed diagonal represents a random classifier (AUC = 0.5). Here, the model's Area Under the Curve (AUC) is 1.000, signifying perfect discrimination every fracture is detected with no healthy wrists falsely flagged.

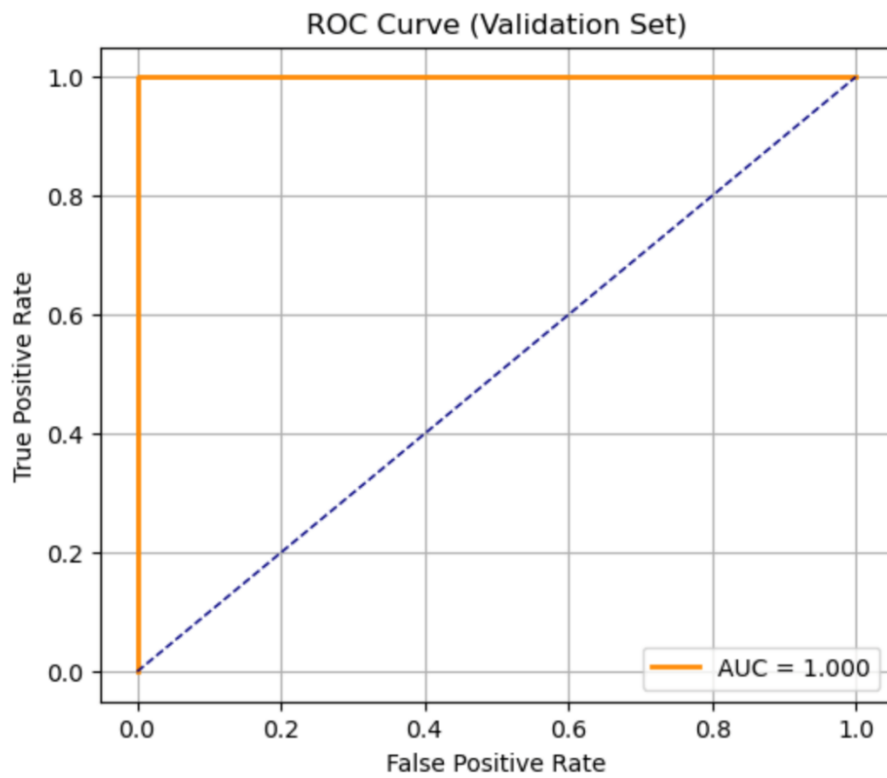


Figure 8. 4. AUCROC of Experiment 1 (Shobayo et al., 2025)

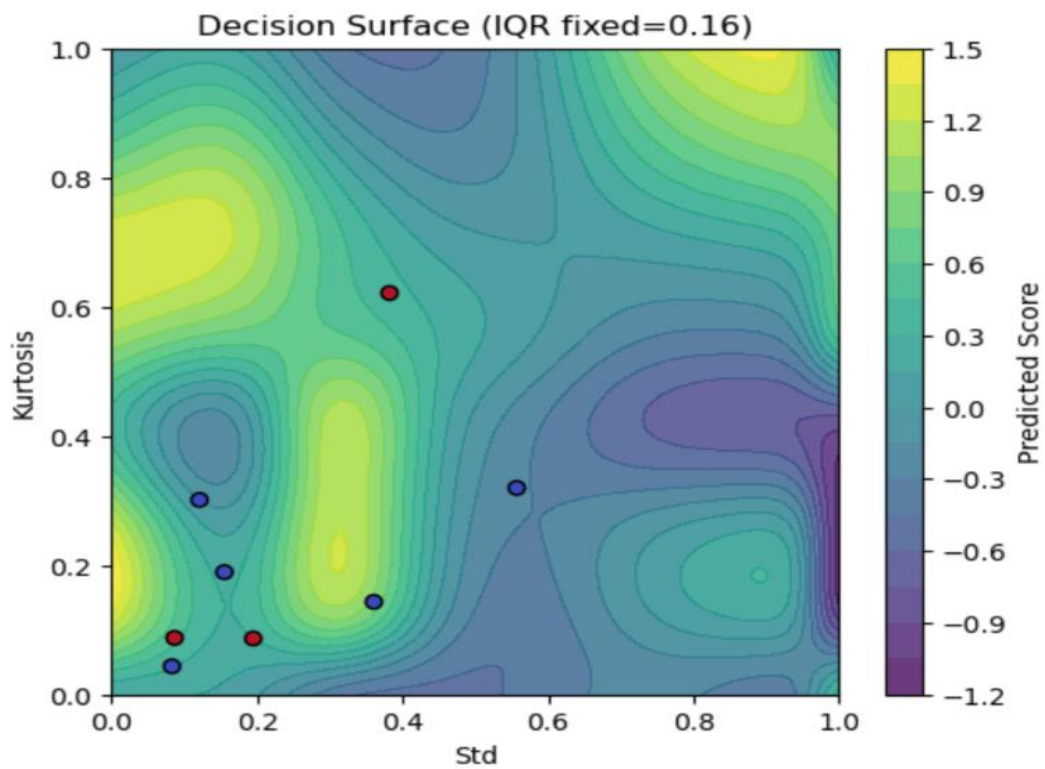


Figure 8. 5. Contour Plot of Experiment 1(Shobayo et al., 2025)

The contour plot in Figure 8.5 visualises the trained ANFIS model’s continuous output (“Predicted Score”) over the 2-D plane of Std (x-axis) and Kurtosis (y-axis), with IQR held fixed at its median value which is approximately 0.16. We have chosen to fix the IQR feature since we have a three-feature model, which will in turn provide a 3-D mapping. Fixing the IQR at the median value will help us able to slice the dimensional viewing of the contour plot in 2-D. The IQR feature was also selected for fixing as it provides the least variations between datapoint from all the 39 subjects used for this study. Selecting the median value provides the best representation of the contour plot prediction for the fracture and non-fracture (sprain) validation dataset. Moving this value up or down changes the shape and position of the yellow and blue regions which in turns places the validation dataset in the wrong contour.

The colour scale in the plot shows the predicted score and the colour bar on the right maps the model’s real-valued output where:

- Yellow/green regions with scores ≥ 0 , indicates stronger “fracture” predictions.
- Blue/purple regions with scores ≤ 0 indicates stronger “no-fracture” (sprain) predictions.

The standard deviation feature of the infrared thermal image (Std) varies from 0 (left) to 1 (right) on the x-axis, why the Kurtosis feature varies from 0 (bottom) to 1 (top) on the y-axis. Every pixel in the plot is a synthetic point, represented by the selected features from the thermal images as evaluated by the ANFIS model.

On the overlaid validation samples, the red dots represent the true fracture cases where $y = 1$ on the validation dataset. The blue dots are the true non-fracture (sprain) cases in which validation dataset are presented as $y = 0$.

Where the background colour shifts toward yellow/green, the model’s score is higher, and these are regions where the model will correctly classify as fracture (threshold is 0.5). Conversely, deep blues and purples are areas the model considers “no fracture” (sprain). It is worth noting that one of the red dots, representing the fracture case land in the bluish zone and this is the only misclassified case from the validation data. The rest of the red dots (fracture cases) lies in the green/yellow areas, representing the correctly predicted fracture cases. The smooth contour lines show how the model interpolates between its learned fuzzy rules. Whenever it peaks, (where there is a bright yellow bump), this corresponds to combinations of Std and Kurtosis where the

model is most confident of a fracture prediction. The valleys (areas representing dark purple troughs) are regions of strong confidence in “no fracture” (sprain) prediction.

8.3.2 Experiment 2: Centre value based on K means clustering for the generalised Membership Function.

For experiment 2, K-means clustering was used to determine the number of clusters for the input data. Internally, K-means iteratively assigns each data point to its nearest cluster and then recomputes each centre as the mean of its assigned points. After convergence, the coordinates of the cluster centroids are stored and sorted so that the membership functions are ascend in the order of their centre location. To ensure that there is a reasonable spread in each Gaussian, we compute the minimum gap between neighbouring centres and set each sigma σ to half that gap. This process ensures the adjacent Gaussians just touch about 0.61 of their peaks, thereby providing good coverage without excessive overlap. During the ANFIS forward pass, each input is converted to its membership grades, effectively creating a data driven, well-spaced Gaussian MFs. By initialising the MFs via K-means rather than randomly, the membership function is aligned with actual clusters in the data.

Table 8. 2. Summary of performance for training and validation targets with epochs(t) for Experiment 2.

Performance Measure	Generalised Membership Function (Generalised bell with K-means centre)				
	Epoch = 800	Epoch = 1600	Epoch = 2400	Epoch = 3200	Epoch = 4000
MSE Training	0.1166	0.1446	0.1185	0.1490	0.1461
MSE Validation	0.1055	0.1121	0.1031	0.1049	0.1056
RMSE Training	0.3414	0.3802	0.3442	0.3860	0.3822
RMSE Validation	0.3248	0.3349	0.3211	0.3238	0.3249
Validation -Training Δ MSE	-0.0110	-0.0324	-0.0154	-0.0441	-0.0405
Elapsed Time (seconds)	9.6	19.3	29.0	38.7	48.4

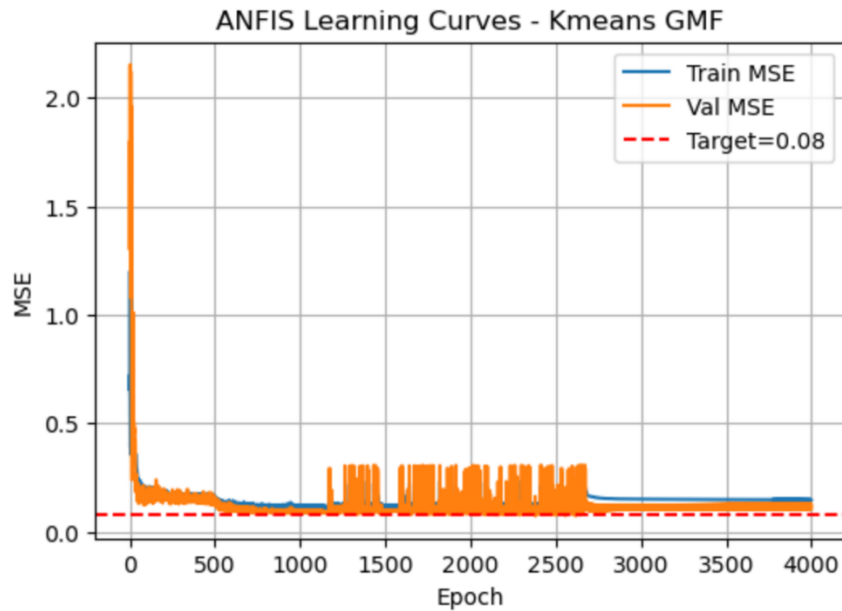


Figure 8. 6. Training and Validation MSE for Experiment 2 (Shobayo et al., 2025)

The Table 8.2 also quantifies error and timing every 800 epochs as in experiment 1. By epoch 800, training MSE is 0.1166 and validation MSE 0.1055 (with validation outperforming training, $\Delta\text{MSE} = -0.0110$). RMSEs mirror these trends (0.341 vs. 0.325). Further snapshots show the MSEs oscillate slightly but stay within 0.10 – 0.15 for both sets, and the ΔMSE remains negative, indicating the model generalises slightly better than it fits the training data. Elapsed time grows linearly at roughly 9.6s per 800 epochs, totalling about 48s at 4,000.

The plot in Figure 8.6 shows the training (blue) and validation (orange) MSE over 4,000 epochs after initializing Gaussian MF centres via K-means. Both curves plunge sharply within the first 200 epochs, reaching low MSE values near the 0.08 target. However, the validation curve then exhibits jagged fluctuations, especially between 1,000 and 2,500 epochs before settling again toward the baseline. The training curve remains relatively smooth.

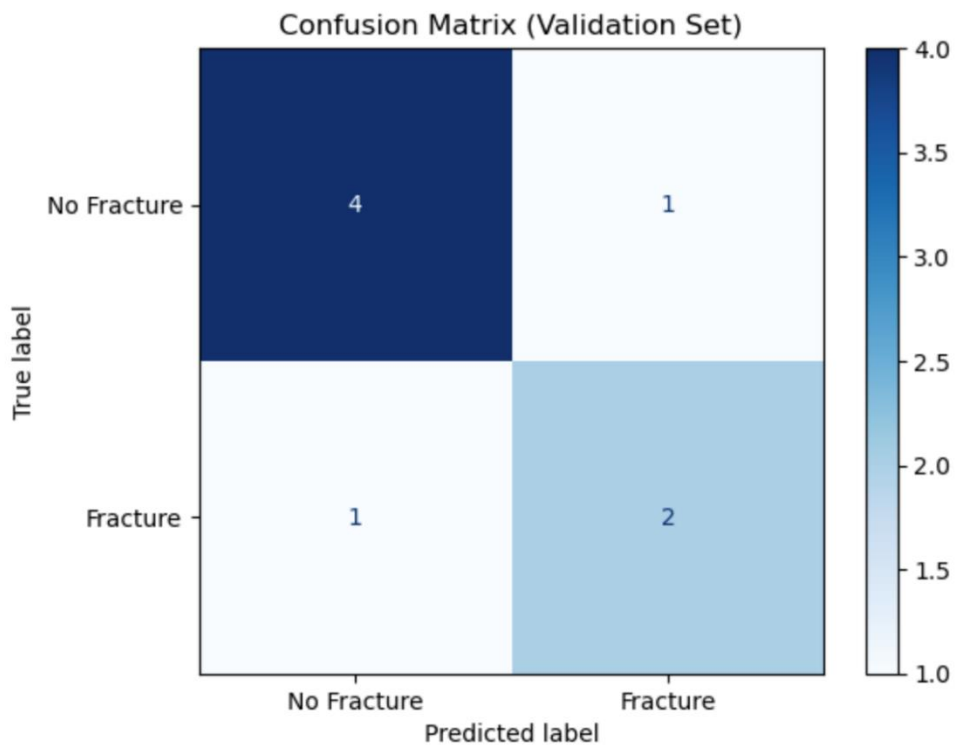


Figure 8. 7. Confusion matrix of Experiment 2 (Shobayo et al., 2025)

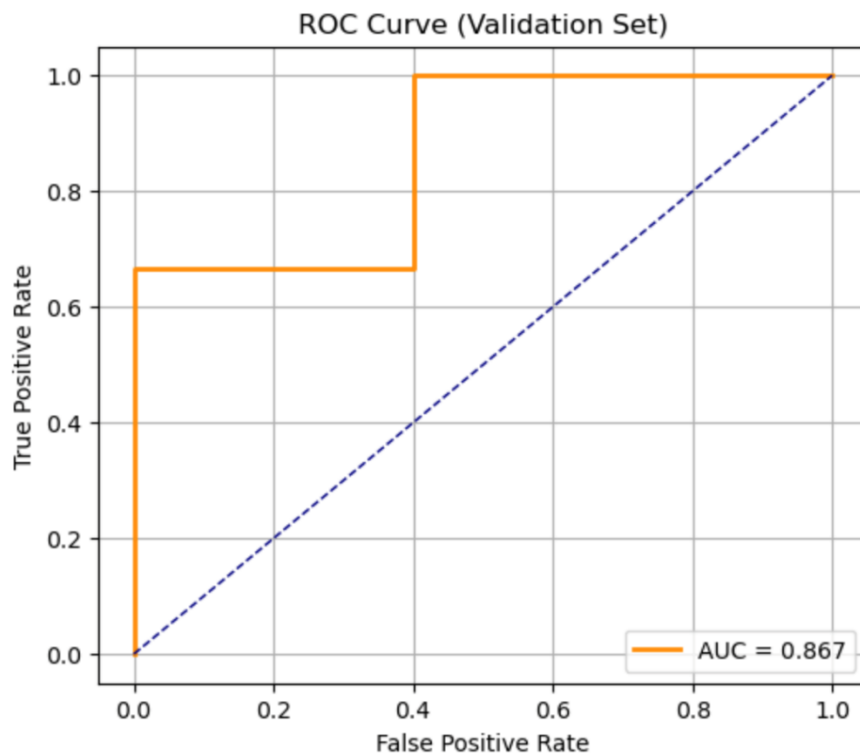


Figure 8. 8. AUCROC for Experiment 2 (Shobayo et al., 2025)

The Figure 8.7 shows the confusion matrix for the K-means initialised ANFIS model on the validation set. Rows show the true class (“No Fracture” on top, “Fracture” below) and columns

show the predicted class. Out of 5 true non-fracture cases, the model correctly labels 4 and misclassifies 1 as fracture (false positive). Of 3 true fractures, it correctly identifies 2 (true positives) but misses 1 (false negative). In total, 6 of 8 cases are classified correctly, yielding an overall accuracy of 75%. The false-positive and false-negative counts also let us compute specificity of 80% and sensitivity of 67%.

The Figure 8.8 shows the ROC curve of the K-means initialised ANFIS. The orange curve rises quickly, where sensitivity jumps to 0.67 with zero false positives, then reaches 1.0 at FPR = 0.5 before flattening at the top right. The Area Under the Curve (AUC) is 0.867, which indicates that there is an 86.7% chance the model ranks a randomly chosen fracture case higher than a non-fracture.

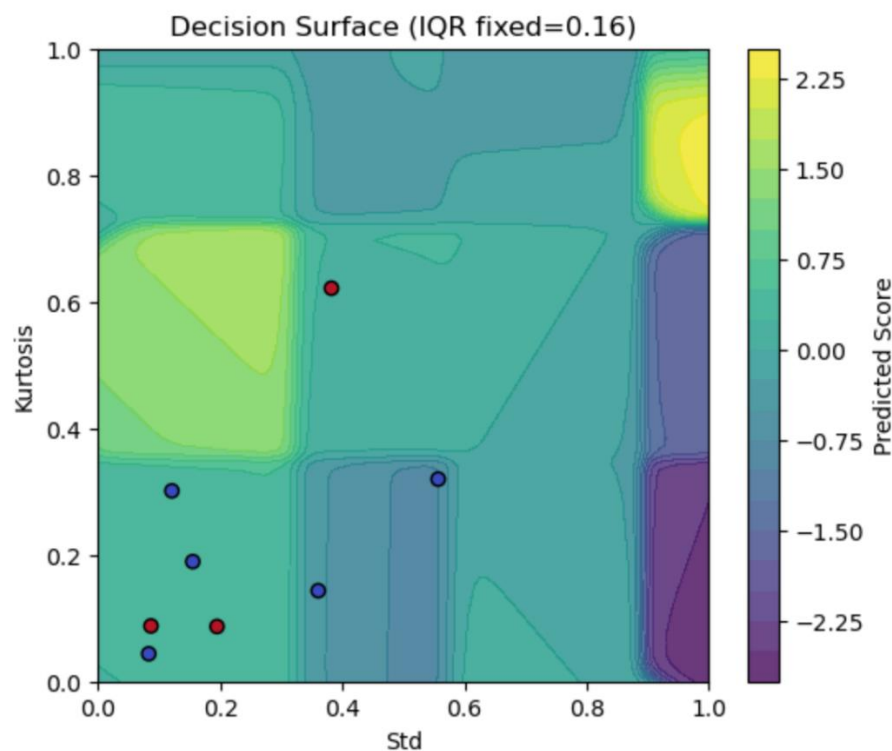


Figure 8. 9. Contour Plot of Experiment 2 (Shobayo et al., 2025)

The Figure 8.9 visualises the K-means initialised ANFIS model’s continuous output over the 2-D plane of Std (x-axis) and Kurtosis (y-axis), with IQR fixed at its median value (≈ 0.16), with each background colour representing the model’s predicted score. Ideally, most red dots which represents the fracture cases lie in yellow regions (correctly predicted), while blue dots (Sprain) fall into cooler areas. Here, a few misclassifications occur where points cross those abrupt rule

boundaries, highlighting areas where the model’s partitioning could be further refined or smoothed.

8.3.3 Experiment 3: Centre value based on fuzzy c-means clustering for the generalised membership function.

For experiment 3, Fuzzy C-means clustering was used to place the gaussian membership function centre based the spatial dimension of the input data. Each input feature i.e., (std, Kurtosis and IQR), a vector of observed values was obtained from the training data and reshaped to fit the fuzzy function parameters. We set the fuzzifier exponent to a value of 3. This helps to control the softness of the cluster boundaries. FCM updates the membership matrix μ_{ik} , for each iteration up to the maximum iteration value ($maxiter = 2000$) or whether the change in centre falls below the error value set at $e = 0.000001$. The centre value obtained for every iteration is then sorted from lowest to highest to represent the Gaussian MF centre of the input. To ensure the Gaussians cover the input range without excessive overlap, the difference between the adjacent centres were computed such that each width is set equal to half the smallest centre-to-centre gap, so the neighbouring Gaussians meet at $e^{-0.5} \approx 0.61$ of their peaks. During prediction, each input value is converted into a membership grade of the gaussian function, ensuring that the membership functions are anchored to actual data clusters rather than arbitrary points.

Table 8. 3. Summary of performance for training and validation targets with epochs(t) for Experiment 3.

Performance Measure	Generalised Membership Function (Gaussian bell with FCM centre)				
	Epoch = 800	Epoch = 1600	Epoch = 2400	Epoch = 3200	Epoch = 4000
MSE Training	0.1890	0.1904	0.1916	0.1768	0.1706
MSE Validation	0.7152	0.5885	0.5212	0.3406	0.1668
RMSE Training	0.4347	0.4363	0.4377	0.4205	0.4130
RMSE Validation	0.8457	0.7672	0.7220	0.5836	0.4084
Validation -Training Δ MSE	0.5262	0.3981	0.3297	0.1638	-0.0038
Elapsed Time (seconds)	1.2	2.4	3.6	4.8	6.0

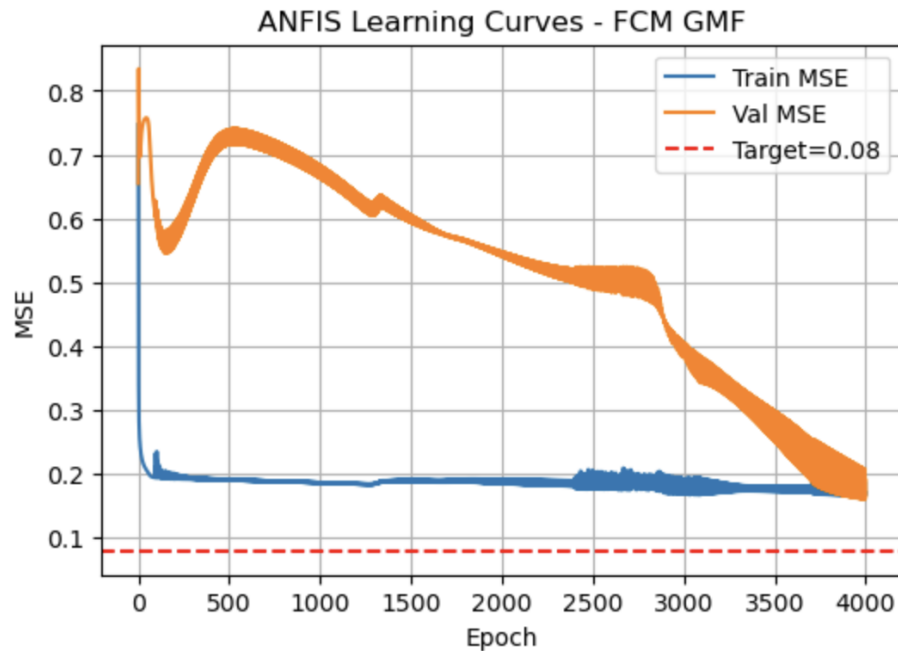


Figure 8.10. Training and Validation MSE for Experiment 3 (Shobayo et al., 2025)

The plot in Figure 8.10 shows how the MSE evolve when Gaussian membership-function centres are initialised by Fuzzy C-Means. Both curves plummet sharply in the first 50 – 100 epochs. The training MSE then stabilises around 0.19, with only minor ripple, while the validation MSE initially peaks near 0.75, then steadily declines to about 0.17 by epoch 4,000 eventually crossing below the training curve.

The Table 8.3 quantifies these same trends at 800-epoch intervals. At epoch 800, training MSE is 0.1890 and validation MSE 0.7152 ($\Delta MSE = 0.5262$). By epoch 2,400, validation error has fallen to 0.5212, closing the gap. By epoch 4,000 both errors converge (training = 0.1706, validation = 0.1668; $\Delta MSE \approx -0.0038$). RMSE values mirror this pattern (training ~ 0.41 , validation dropping from ~ 0.85 to ~ 0.41). The total training time is significantly low at only 6 s for 4,000 epochs.

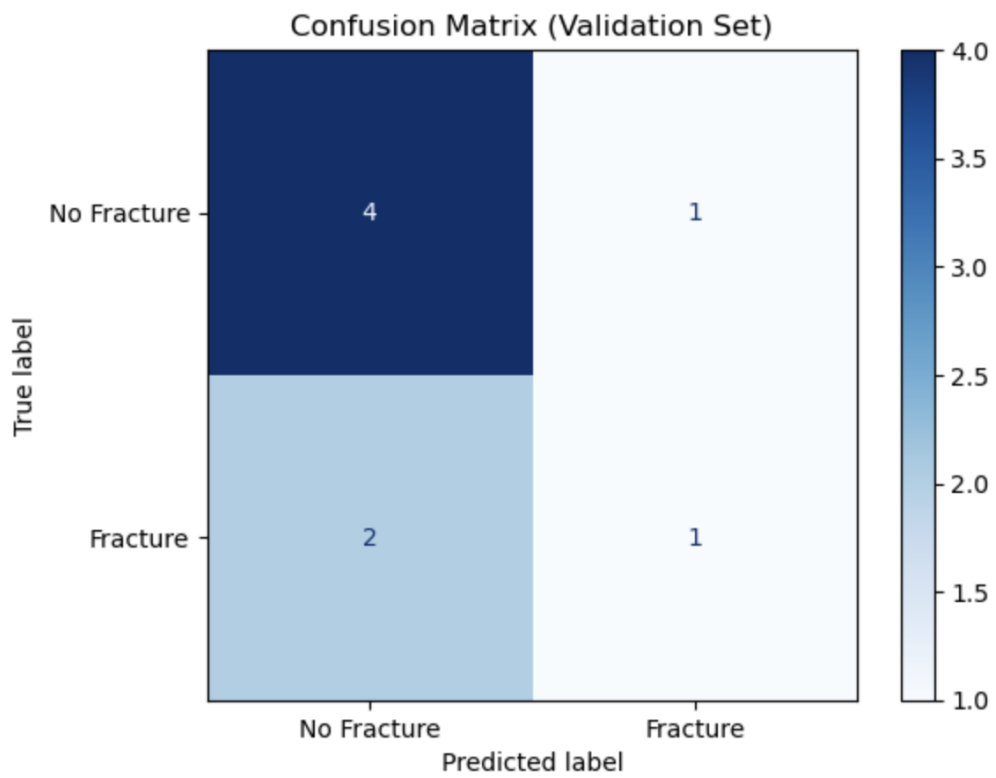


Figure 8. 11 Confusion Matrix for Experiment 3 (Shobayo et al., 2025)

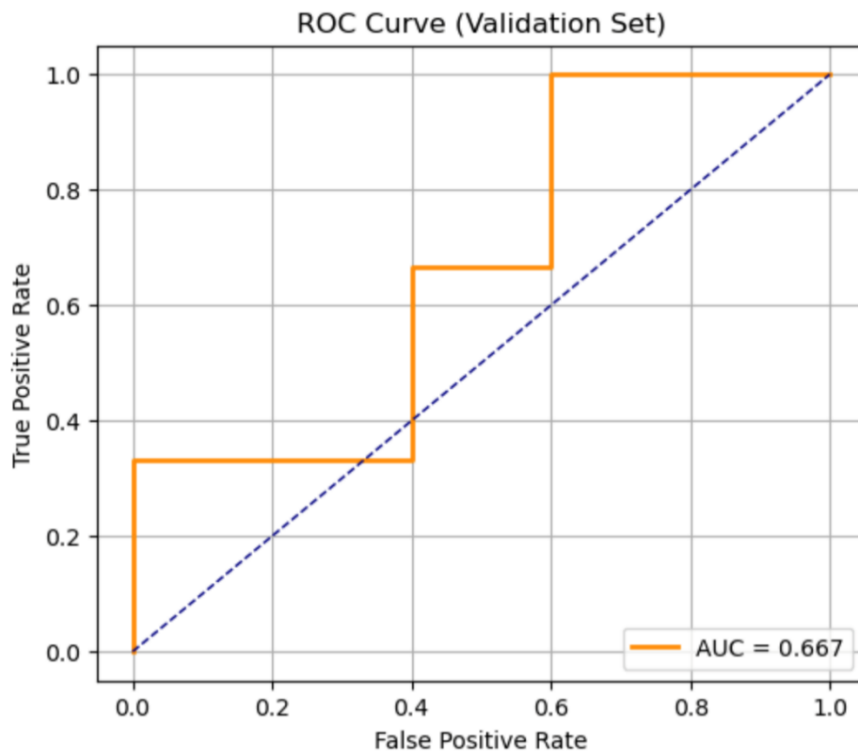


Figure 8. 12. AUCROC for Experiment 3 (Shobayo et al., 2025)

The Figure 8.11 shows the confusion matrix for the FCM-initialised ANFIS on the validation set. Out of five true non-fracture cases, four are correctly identified (true negatives) and one is

misclassified as fracture (false positive). Out of three true fractures, only one is correctly detected (true positive) while two are missed (false negatives). This yields an overall accuracy of 62.5%, specificity of 80%, and sensitivity of 33%.

The Figure 8.12 is ROC curve for the FCM-initialised ANFIS, showing how sensitivity varies with the false-positive rate as the classification threshold shifts. The orange curve initially jumps to 0.33 true positive rate at zero false positives, then to 0.67 at FPR = 0.4, and finally reaches 1.0 at FPR = 1.0. The resulting AUC of 0.667 indicates modest discriminative power: the model is better than chance but still struggles to reliably distinguish fractures from non-fractures under this FCM initialisation.

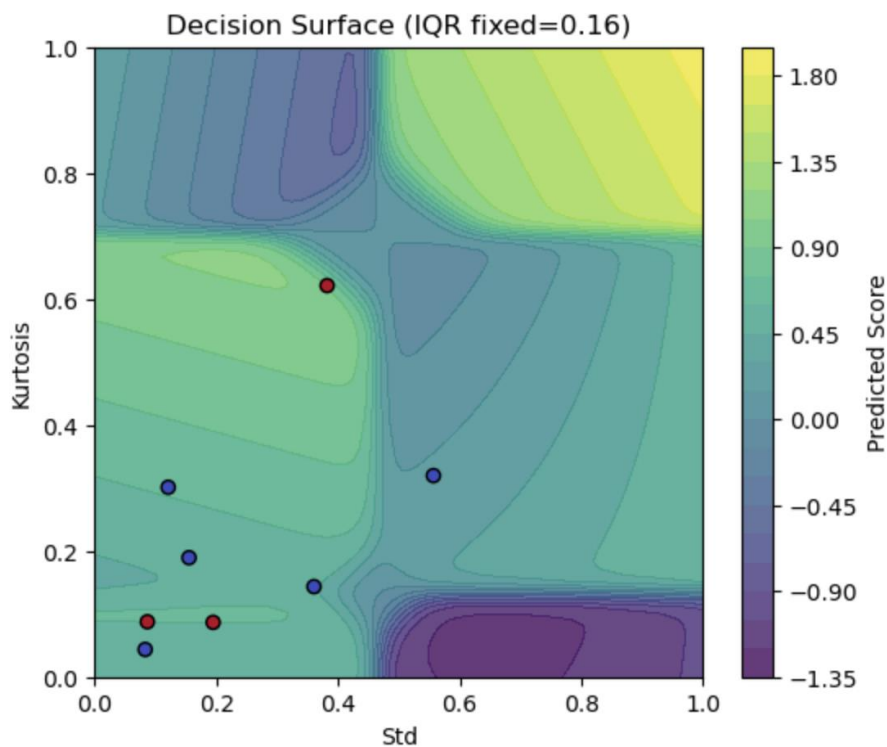
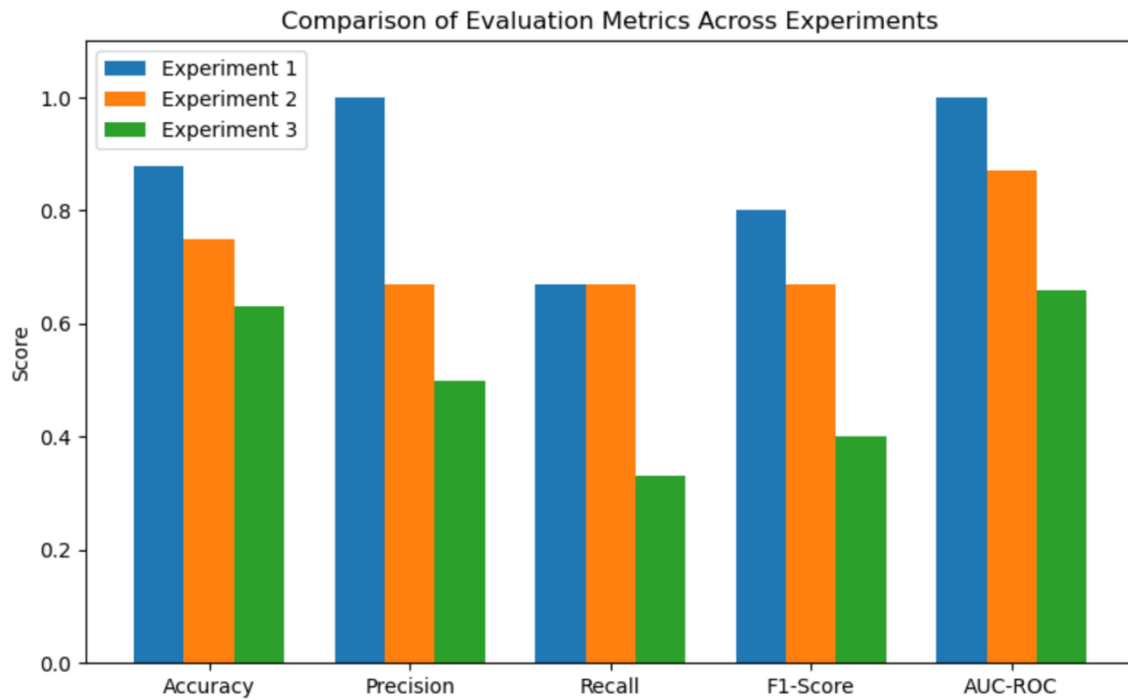


Figure 8. 13. Contour plot for Experiment 3 (Shobayo et al., 2025)

This Figure 8.13 shows the FCM-initialised ANFIS decision surface over two features i.e., Std (x-axis) and Kurtosis (y-axis), with IQR fixed at 0.16. Ideally, red points lie in yellow regions and blue in purple. Here, most red dots fall in positive (yellow/green) zones, but one lies in a cooler region, showing a misclassification. The irregular contours highlight how FCM-initialized Gaussians partition the feature space for fracture detection. The comparison of the evaluation metrics of the 3 experiments is shown in Table 8.4 and Figure 8.14

Table 8. 4. Comparison of evaluation metrics

Evaluation Metrics	Experiment 1	Experiment 2	Experiment 3
Accuracy	0.88	0.75	0.63
Precision	1.00	0.67	0.50
Recall	0.67	0.67	0.33
F1-Score	0.80	0.67	0.40
AUC-ROC	1.00	0.87	0.66

**Figure 8. 14.** Evaluation metrics across the experiments (Shobayo et al., 2025)

Across the three experiments, there was a clear trade-off between the convergence speed, generalisation and classification accuracy. Experiment 1 (random centres) achieved steady but relatively slow MSE declines, requiring nearly 194s to complete 4000 epochs and settling at a training/validation MSE of 0.1304/0.1514. Its Δ MSE narrowed from 0.039 to 0.021, indicating modest overfitting. Experiment 2 (K-means) converged faster, taking only 48s for 4000 epochs, with training/validation MSE fluctuating around 0.14/0.10 and Δ MSE remaining slightly negative (-0.011 to -0.041), showing slight underfitting but excellent stability. Experiment 3 (FCM) was fastest (6 s for 4000 epochs) but exhibited unstable validation error, initially peaking near 0.75 before gradually falling to 0.167, finally matching training MSE at 0.171 (Δ MSE \approx -0.004). The dramatic early spikes in Experiment 3 suggest poor initial cluster placement.

Random initialization (Experiment 1) can achieve high accuracy but at the expense of longer training time and slight overfitting. K-means (Experiment 2) strikes the best balance: it leverages

data-driven cluster placement to accelerate convergence, reduce overfitting, and yield robust, reproducible results. FCM (Experiment 3) offers the quickest training but suffers from early instability and poor discrimination, likely because its fuzzy assignments blur crucial rule partitions. Overall, K-means-driven ANFIS emerges as the most effective strategy, combining computational efficiency with strong and consistent classification performance.

The accuracy of prediction was compared with the results in chapter 6 (MLP) and chapter 7 (CNN) and Experiment 1 with the generalised bell MF had the highest accuracy which using the same dataset as shown in Figure 8.15. K-means centred cluster was very close in terms of accuracy but proffer faster training convergence when compared to the models with higher accuracy.

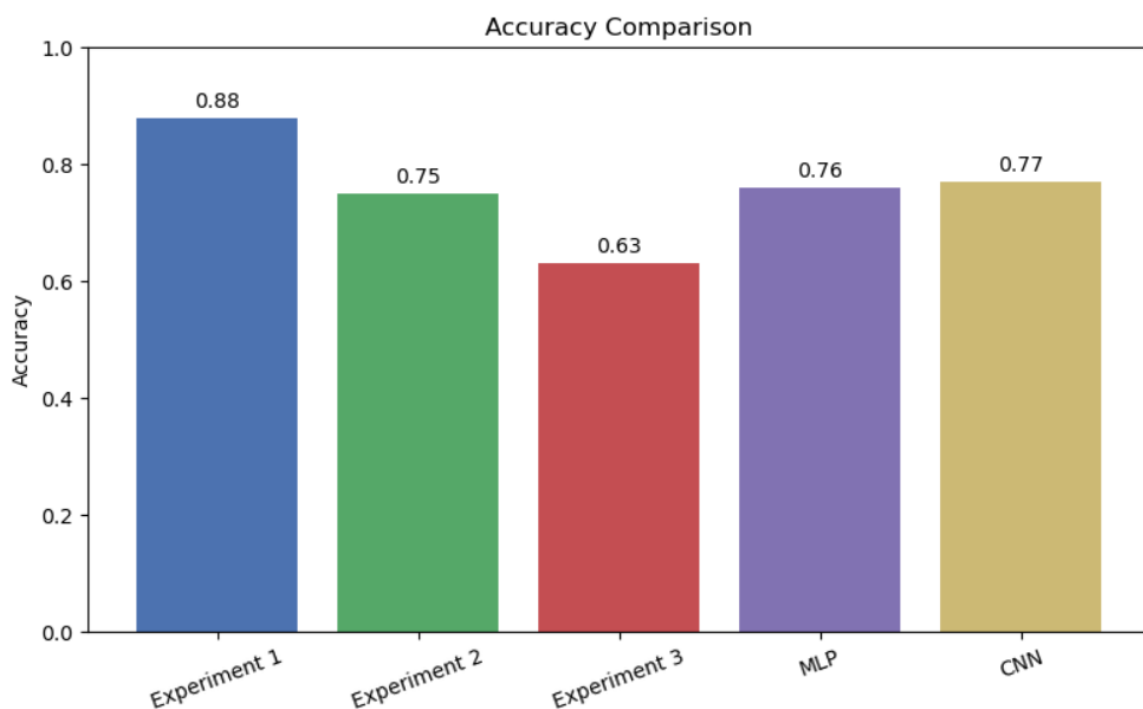


Figure 8. 15 Accuracy comparison with other models (Shobayo et al., 2025)

8.4 Summary

This study demonstrated that ANFIS can effectively discriminate paediatric wrist fractures from sprains using a compact set of thermographic features (standard deviation, inter quartile range, kurtosis) derived from infrared recordings. By embedding Takagi – Sugeno fuzzy rules within a trainable network, the approach overcomes the opacity of conventional “black box” classifiers while retaining nonlinear modelling capacity, allowing each learned rule and membership function to be clinically scrutinised and, if needed, refined by domain experts. A central finding was the influence of premise parameter initialisation on both convergence dynamics and

generalisation: K-means seeded Gaussian memberships yielded the most favourable balance, which included fast optimisation with stable validation error, surpassing the slower yet high precision random seeding and the rapid but volatile fuzzy C means alternative (higher early validation instability, lower AUC). This indicates that data aware, crisp clustering can establish sufficiently separated initial fuzzy partitions without inducing the excessive overlap that may dilute discriminative rule firing seen in soft clustering schemes. The hybrid least square and gradient descent training further reduced search space dimensionality, contributing to efficient error minimisation and supporting deployment within time sensitive emergency workflows.

Chapter 9 – Conclusion and Further work

9.1 Conclusion

In this study, artificial intelligence neural network-based approaches were used to classify infrared thermal images of wrist fracture and sprain of paediatric. The study included 40 participants, 24 males and 16 females, mean age 10.50 years (standard deviation 2.63 years), 19 with a wrist fracture and 21 with a sprain, with the injury type confirmed by x-ray radiography. The mean body temperature was 36.3 °C (standard deviation 0.43 °C) across all participants. The developed techniques were used to classify wrists' infrared thermal images and compare the diagnostic results with the x-ray radiography to develop an automated diagnostic tool for fracture detection.

To achieve this aim, the following objectives were met:

- i. A suitable infrared thermal imaging preprocessing and feature extraction technique to help in the differentiation of wrist fracture from sprain was developed.
- ii. The correlation between time since injury and skin temperature for fracture and sprain was examined.
- iii. A novel feature extraction method was devised and with multilayer perceptron (MLP) neural network classified fracture and sprain.
- iv. The effectiveness of convolutional neural network to classify the infrared thermal images was compared with the manual feature extraction methods utilised in MLP.
- v. To introduce explainability in decision making, a fuzzy-based ANFIS model to discriminate between the infrared thermal images of fracture and sprain was devised and its effectiveness analysed.

9.2 Study's findings and recommendations

The findings and recommendations of this study are listed in the following headings.

9.2.1 Correlation of skin temperature and time since injury

The skin temperature increase at the injury site had a lower standard deviation around its mean value for fracture as compared to sprain. For fracture, two phases with distinct temperature profiles were observed. For the initial phase, lasting up to around 3.3 hours after injury, there was no significant correlation between time since injury (TSI) and change in temperature (ΔT) but thereafter a moderate correlation was observed. For sprain, ΔT did not correlate significantly with

TSI. These findings may relate to differences in the inflammatory and healing processes between fractures and sprains. The correlation difference between the first and second phase observed in the ΔT versus TSI plot, may be indicative of the manner fracture healing affects skin temperature at the injury site. To be able to more accurately establish the boundary between the two phases for the ΔT -TSI plot in fracture, a larger sample size with a more uniform TSI would be required. Previous studies indicated that the skin temperature increase at the site of fracture subsides as the healing progresses. Given the sample size, this phenomenon was not explored in this study. This study used a linear correlation model to explore ΔT versus TSI relationship. For a larger sample size, a higher order polynomial could be used to detect finer details.

9.2.2 Multilayer perceptron based IRT image classification model

The ability of an MLP in differentiation tasks is affected by the manner of data pre-processing and feature extraction. The grid representation of the ROI for injured wrists allowed pixels spatially close to each other with similar temperature values to be expressed by their mean values. This then allowed the 50 largest mean temperature values from the injured wrist ROI to be selected for characterisation by statistical measures. To deal with skin temperature variability across the participants, the mean temperature of the contralateral (uninjured) ROI was subtracted from the selected values. Each cell within the grid was represented by 10×10 pixels, i.e., total 100 pixels. The cell dimension was determined by experimenting with different sizes and observing the MLP differentiation outcomes. In future studies, a more in-depth analysis could be carried out in determining the cell size. The statistical measures used to analyse the 50 selected values, quantified the magnitude of the temperature (i.e., maximum, minimum, mean, mode and median), temperature deviation from their mean value, interquartile range and distribution (i.e., skewness and kurtosis). Standard deviation from the mean, kurtosis, and interquartile range (IQR) proved more effective for differentiating between the two types of injuries and thus were used as input to the MLP. The selection of these three measures should be considered in the context of the limited number of participants included in this study, and thus, with a larger number of participants, the effectiveness of these three measures needs further evaluation. The manner of randomly selecting participants for MLP training and test files and averaging the MLP outputs over 100 trials was to mitigate the bias that can occur when selecting patients for each file. This bias may alter the results when the sample size is not large. MLP can differentiate between nonlinearly separable data and is not sensitive to the type of data distribution. These can make MLP a more robust classifier than statistical methods such as linear discriminant analysis. The discrimination sensitivity (84.2%) and specificity (71.4%) indicated that the method was more

effective in correctly identifying the participants with fracture (16 participants from 19) than correctly identifying participants with sprain (15 participants from 21). Comparison of the results from investigations A and B indicated that changes in the number of participants in the training and test files had not affected the sensitivity and specificity. The severity of injury types (fracture and sprain) varied across the participants. It is possible that a severe sprain may cause a greater temperature increase at the site of the injury than a small fracture. False negatives (fracture classified as sprain) have a more serious impact on treating patients than false positives (sprain classified as fracture). There were three false negatives and six false positive cases.

9.2.3 Convolutional neural network based IRT image classification model

The purpose of this study was to develop an IRT image classification model utilising a CNN-based deep neural network, using the automatic feature selection, when compared to the manual process used in the MLP algorithm. The model produced an overall accuracy of 76% for the classification of the IRT images between wrist fracture and non-fracture (sprain) when image augmentation was applied without dropout. This provided improved accuracy compared to when a dropout layer was introduced without augmentation. The model also showed an NPV of 80% and a PPV of 73%. Therefore, the classification of images associated with the non-fractured wrists was 80% correct. The ROC curve presented indicated the model's operating point of 0.82 which provided a better measure of classification considering a larger number of non-fractured (sprain) images in the dataset. The ROC area under the curve (AUC) was used to further highlight the ability of the model in differentiating the injury types. It provided a benchmark for clinical diagnosis regardless of the number of classes. When the model was used without augmentation and with a dropout value of 0.2, the model produced a specificity of 68% and a sensitivity of 50%. These values were much lower compared to the first experiment. This can be attributed to the nature of the IRT images used in the CNN architecture. The original images had small variations, and dropping out played a significant role in the model accuracy as pattern recognition became harder. The inclusion of the image augmentation technique increased the variability of the IRT dataset, assisting CNN generalisation during its training.

9.2.4 Adaptive Neuro-fuzzy Inference system based IRT image classification model

By assessing an adaptive neuro-fuzzy system (ANFIS) on the IRT images across the three experiments, there was a clear trade-off between convergence speed, generalisation and classification accuracy. Experiment 1 (random centres) achieved steady but relatively slow MSE declines, requiring nearly 194 s to complete 4000 epochs and settling at a training/validation MSE of 0.130/0.151. Its Δ MSE narrowed from 0.039 to 0.021, indicating modest overfitting.

Experiment 2 (K-means) converged faster, taking only 48 s for 4000 epochs, with training/validation MSE fluctuating around 0.14/0.10 and Δ MSE remaining slightly negative (-0.0110 to -0.0405), showing slight underfitting but excellent stability. Experiment 3 (FCM) was fastest (6 s for 4000 epochs) but exhibited unstable validation error, initially peaking near 0.75 before gradually falling to 0.167, finally matching training MSE at 0.171 (Δ MSE \approx -0.004). The dramatic early spikes in Experiment 3 suggest a poor initial cluster placement. Experiment 1 delivered high specificity and perfect precision i.e., Accuracy 0.88, Precision 1.00, Recall 0.67, F1 0.80, AUC 1.00, thus reflecting near-ideal separation on the small validation set. Experiment 2 traded precision for broader generalisation (Accuracy 0.75, Precision 0.67, Recall 0.67, F1 0.67, AUC 0.87), yet provided the most balanced sensitivity and specificity. For Experiment 3, the late-stage validation improvements were insufficient to overcome early misclassifications (Accuracy 0.63, Precision 0.50, Recall 0.33, F1 0.40, AUC 0.67), indicating that its soft clustering induced ambiguous rule boundaries.

Random initialisation (Experiment 1) can achieve high accuracy but at the expense of longer training time and slight overfitting. K-means (Experiment 2) strikes the best balance: it leverages data-driven cluster placement to accelerate convergence, reduce overfitting, and yield robust, reproducible results. FCM (Experiment 3) offers the quickest training but suffers from early instability and poor discrimination, likely because its fuzzy assignments blur crucial rule partitions.

9.2.5 Limitations of the study

The possible limitations to this study include the small sample size, injury-timing variability which could influence the temperature measurement during data collection and the semblance of the IRT image datasets, making it difficult to be trained using transfer learning CNN and Transformer models.

9.3 Future work

Building on the proof-of-concept studies presented in this work, future efforts should first enlarge sample size and diversity via multi-centre, longitudinal trials spanning paediatric subgroups, adult cohorts, and multiple fracture sites (e.g., ankle, humerus, hip). Broader sampling will validate temperature-injury relationships across demographics, refine temporal-phase models of inflammation and repair, and quantify inter- and intra-subject variability under standardised acquisition (emissivity control, camera distance/angle) with blackbody or reference targets for calibration. Integrating real-time infrared thermal imaging with wearable sensors (e.g., skin thermistors, photoplethysmography, accelerometers) could capture continuous perfusion

dynamics and movement context, linking cutaneous temperature fluctuations to deeper haemodynamic and healing markers.

For the machine learning pipeline, assembling larger, class-balanced datasets will support richer architectures: transfer-learning CNNs pretrained on medical imagery; graph neural networks that encode spatial and bilateral ROI relationships; and hybrid models that fuse MLP, CNN, and ANFIS modules to balance accuracy with interpretability. Generative models such as GANs and VAEs could be developed to synthesise realistic thermal sequences for minority conditions, enabling few-shot learning and stress-testing robustness. Domain adaptation, harmonisation across devices or sites, and rigorous external validation are essential. Uncertainty estimation and calibration should accompany predictions to support risk-aware clinical decisions.

Explainable AI ought to be embedded end-to-end. For example, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) for feature attributions, Grad-CAM (Gradient-weighted Class Activation Mapping) for spatial saliency, and counterfactual analyses to help understand surface model rationales, reveal failure modes, and foster clinician trust through human-in-the-loop review. Finally, translating these workflows to mobile or point-of-care platforms, with on-device inference, privacy-preserving learning (e.g., federated approaches), and seamless EHR (Electronic Health Record) integration will enable rapid, non-invasive triage in emergency and remote settings, while usability, regulatory, and cost-effectiveness studies will ensure safe, equitable deployment.

References

- Abakka, M., Ouagari, H. E., Baidriss, Y., Mekkaoui, J., Bouffetal, M., Bassir, R., Kharmaze, M., Lamrani, M. O., & Berrada, M. S. (2023). *The Treatment of Comminuted Fractures of the Distal Extremity of the Radius by External Fixator: About 12 Cases*. AMO Publisher. 10.59324/Ejtas.2023.1(3).48
- Abbaoui, W., Retal, S., Ziti, S., & El Bhiri, B. (2024). *Automated Ischemic Stroke Classification from MRI Scans: Using A Vision Transformer Approach*. MDPI AG. 10.3390/Jcm13082323
- Abdulahi, A. T., Ogundokun, R. O., Adenike, A. R., Shah, M. A., & Ahmed, Y. K. (2024). *Pulmonet: A Novel Deep Learning Based Pulmonary Diseases Detection Model*. Springer Science and Business Media LLC. 10.1186/S12880-024-01227-2
- Ahmad, W., Ali, H., Shah, Z., & Azmat, S. (2022). A New Generative Adversarial Network for Medical Images Super Resolution. *Scientific Reports*, 12(1)10.1038/S41598-022-13658-4
- Al-Ali, A., Elharrouss, O., Qidwai, U., & Al-Maaddeed, S. (2021). *ANFIS-Net for Automatic Detection Of COVID-19*. Springer Science and Business Media LLC. 10.1038/S41598-021-96601-3
- Alam, A., Al-Shamayleh, A. S., Thalji, N., Raza, A., Morales Barajas, E. A., Thompson, E. B., De La Torre Diez, I., & Ashraf, I. (2025). Novel Transfer Learning Based Bone Fracture Detection Using Radiographic Images. *BMC Medical Imaging*, 25(1), 5.
- Alomar, K., Aysel, H. I., & Cai, X. (2023). Data Augmentation in Classification and Segmentation: A Survey and New Strategies. *Journal Of Imaging*, 9(2), 46.
- Alzubaidi, L., Al-Amidie, M., Al-Asadi, A., Humaidi, A. J., Al-Shamma, O., Fadhel, M. A., Zhang, J., Santamaría, J., & Duan, Y. (2021). *Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data*. MDPI AG. 10.3390/Cancers13071590
- Amarneni, S., & Valarmathi, D. R. S. (2024). Diagnosing The MRI Brain Tumour Images Through RNN-LSTM. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, 910.1016/J.Prime.2024.100723
- Anbalagan, V., & Balasubramanian, V. (2024). HBO-GMRNN: Honey Badger Optimization Based Gain Modulated Recurrent Neural Network for Classification of Breast Cancer. *Biomedical Signal Processing and Control*, 9110.1016/J.Bspc.2023.105910

- Antamis, T., Drosou, A., Vafeiadis, T., Nizamis, A., Ioannidis, D., & Tzovaras, D. (2024). Interpretability Of Deep Neural Networks: A Review of Methods, Classification and Hardware. *Neurocomputing*, *601*, 128204.
- Arthur, D. T., Khan, M. M., & Barclay, L. C. (2011). Thermographic Investigation of Osseous Stress Pathology. Paper Presented at the *2011 Annual International Conference of The IEEE Engineering in Medicine and Biology Society*, 6250–6253.
- Atasever, S., Azginoglu, N., Terzi, D. S., & Terzi, R. (2022). A Comprehensive Survey of Deep Learning Research on Medical Image Analysis with Focus on Transfer Learning. *Clinical Imaging*, *94*, 18. 10.1016/J.Clinimag.2022.11.003
- Auten, J. D., Naheedy, J. H., Hurst, N. D., Pennock, A. T., Hollenbach, K. A., & Kanegaye, J. T. (2019). Comparison of Paediatric Post-Reduction Fluoroscopic-And Ultrasound Forearm Fracture Images. *The American Journal of Emergency Medicine*, *37*(5), 832–838.
- Avci, E., & Turkoglu, I. (2009). An Intelligent Diagnosis System Based on Principle Component Analysis and ANFIS for the Heart Valve Diseases. *Expert Systems with Applications*, *36*(2), 2873–2878.
- Avdelidis, N. P., & Almond, D. P. (2004). Transient Thermography as a Through Skin Imaging Technique for Aircraft Assembly: Modelling and Experimental Results. *Infrared Physics & Technology*, *45*(2), 103–114.
- Avdelidis, N. P., Gan, T., Ibarra-Castanedo, C., & Maldague, X. (2011). Infrared Thermography as a Nondestructive Tool for Materials Characterisation and Assessment. Paper Presented at the *Thermosense: Thermal Infrared Applications XXXIII*, 8013 308–314.
- Ayana, G., Dese, K., Abagaro, A. M., Jeong, K. C., Yoon, S., & Choe, S. (2024). Multistage Transfer Learning for Medical Images. *Artificial Intelligence Review*, *57*(9)10.1007/S10462-024-10855-7
- Ayub, S., Kannan, R. J., Alsini, R., Hasanin, T., & Sasidhar, C. (2022). LSTM-Based RNN Framework to Remove Motion Artifacts in Dynamic Multicontrast MR Images with Registration Model. *Wireless Communications and Mobile Computing*, *2022*, 1. 10.1155/2022/5906877
- Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., Karimijafarbigloo, S., Cohen, J. P., Adeli, E., & Merhof, D. (2024). Medical Image Segmentation Review: The Success Of U-Net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

- Azeez, O., & Abdulazeez, A. (2025). Classification Of Brain Tumor Based on Machine Learning Algorithms: A Review. *Journal Of Applied Science and Technology Trends*, 6(1), 1.
- B., A., & Kalirajan, K. (2024). An Intelligent Magnetic Resonance Imaging-Based Multistage Alzheimer's Disease Classification Using Swish-Convolutional Neural Networks. *Medical & Biological Engineering & Computing*, 10.1007/S11517-024-03237-2
- Balachandar, N., Chang, K., Kalpathy-Cramer, J., & Rubin, D. L. (2020). Accounting For Data Variability in Multi-Institutional Distributed Deep Learning for Medical Imaging. *Journal of the American Medical Informatics Association*, 27(5), 700–708.
- Balakrishnan, N. D., & Perumal, S. K. (2025). Monitoring Kidney Microanatomy During Ischemia-Reperfusion Using ANFIS Optimized CNN. *International Urology and Nephrology*, , 1–12.
- Balaras, C. A., & Argiriou, A. A. (2002). Infrared Thermography for Building Diagnostics. *Energy And Buildings*, 34(2), 171–183.
- Baldi, P. (2012). Autoencoders, Unsupervised Learning, And Deep Architectures. Paper Presented at the *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, 37–49.
- Bannerman, S. L., Lim, C., Moon, J., Nicholls, J., Straight, P., & Thompson, M. Treatment of Sprains and Strains. *Nzms J*, 27.
- Baoge, L., Van Den Steen, E., Rimbaut, S., Philips, N., Witvrouw, E., Almqvist, K. F., Vanderstraeten, G., & Vanden Bossche, L. C. (2012). Treatment Of Skeletal Muscle Injury: A Review. *International Scholarly Research Notices*, 2012(1), 689012.
- Bayer, M., Kaufhold, M., & Reuter, C. (2022). A Survey on Data Augmentation for Text Classification. *ACM Computing Surveys*, 55(7), 1–39.
- Berroukham, A., Housni, K., & Lahraichi, M. (2023-12-16). Vision Transformers: A Review of Architecture, Applications, And Future Directions. Paper Presented at the 205. 10.1109/Cist56084.2023.10410015
- Birgani, M. T., Chegeni, N., Birgani, F. F., Fatehi, D., Akbarizadeh, G., & Shams, A. (2019). Optimization Of Brain Tumor MR Image Classification Accuracy Using Optimal Threshold, PCA And Training ANFIS With Different Repetitions. *Journal Of Biomedical Physics & Engineering*, 9(2), 189.

- Bishop, C. M. (1994). Neural Networks and Their Applications. *Review Of Scientific Instruments*, 65(6), 1803–1832. 10.1063/1.1144830
- Bora, A., & Cuayáhuitl, H. (2024). Systematic Analysis of Retrieval-Augmented Generation-Based Llms for Medical Chatbot Applications. *Machine Learning and Knowledge Extraction*, 6(4), 2355. 10.3390/Make6040116
- Brophy, E., Wang, Z., She, Q., & Ward, T. (2023). Generative Adversarial Networks in Time Series: A Systematic Literature Review. *ACM Computing Surveys*, 55(10), 1–31.
- Buchberger, B., Scholl, K., Krabbe, L., Spiller, L., & Lux, B. (2022). Radiation Exposure by Medical X-Ray Applications. *German Medical Science*, 20, Doc06. 10.3205/000308
- Canavese, F., Samba, A., & Rousset, M. (2016). Pathological Fractures in Children: Diagnosis and Treatment Options. *Orthopaedics & Traumatology: Surgery & Research*, 102(1), S149. 10.1016/J.Otsr.2015.05.010
- Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics*, 8(8), 832.
- Chai. (2011). Gray-Level Co-Occurrence Matrix Bone Fracture Detection. *American Journal of Applied Sciences*, 8(1), 26–32. 10.3844/Ajassp.2011.26.32
- Chai, H. Y., Wee, L. K., Swee, T. T., Salleh, S., & Ariff, A. K. (2011). Gray-Level Co-Occurrence Matrix Bone Fracture Detection. *American Journal of Applied Sciences*, 8(1), 26.
- Chatterjee, S., & Das, A. (2020a). A Novel Systematic Approach to Diagnose Brain Tumour Using Integrated Type-II Fuzzy Logic and ANFIS (Adaptive Neuro-Fuzzy Inference System) Model. *Soft Computing*, 24(15), 11731–11754.
- Chatterjee, S., & Das, A. (2020b). A Novel Systematic Approach to Diagnose Brain Tumour Using Integrated Type-II Fuzzy Logic and ANFIS (Adaptive Neuro-Fuzzy Inference System) Model. *Soft Computing-A Fusion of Foundations, Methodologies & Applications*, 24(15)
- Chau, M. (2024). Ethical, Legal, And Regulatory Landscape of Artificial Intelligence in Australian Healthcare and Ethical Integration in Radiography: A Narrative Review. *Elsevier BV*. 10.1016/J.Jmir.2024.101733
- Chauhan, S., Edla, D. R., Boddu, V., Rao, M. J., Cheruku, R., Nayak, S. R., Martha, S., Lavanya, K., & Nigat, T. D. (2024). Detection Of COVID-19 Using Edge Devices by a Light-Weight

- Convolutional Neural Network from Chest X-Ray Images. *BMC Medical Imaging*, 24(1)10.1186/S12880-023-01155-7
- Chen, C., Zhang, P., Zhang, H., Dai, J., Yi, Y., Zhang, H., & Zhang, Y. (2020). Deep Learning on Computational-Resource-Limited Platforms: A Survey. *Wiley*. 10.1155/2020/8454327
- Chen, I. Y., Pierson, E., Rose, S., Joshi, S., Ferryman, K., & Ghassemi, M. (2021). Ethical Machine Learning in Healthcare. *Annual Review of Biomedical Data Science*, 4(1), 123–144.
- Chen, J., Frey, E. C., He, Y., Segars, W. P., Li, Y., & Du, Y. (2022). Transmorph: Transformer for Unsupervised Medical Image Registration. *Medical Image Analysis*, 8210.1016/J.Media.2022.102615
- Chen, R., Song, Y., Huang, J., Wang, J., Sun, H., & Wang, H. (2021). Rapid Diagnosis and Continuous Monitoring of Intracerebral Haemorrhage with Magnetic Induction Tomography Based on Stacked Autoencoder. *AIP Publishing*. 10.1063/5.0050171
- Chen, Y., Yang, X., Wei, Z., Heidari, A. A., Zheng, N., Li, Z., Chen, H., Hu, H., Zhou, Q., & Guan, Q. (2022). Generative Adversarial Networks in Medical Image Augmentation: A Review. *Computers In Biology and Medicine*, 144, 105382.
- Chetty, G., Yamin, M., & White, M. (2022). A Low Resource 3D U-Net Based Deep Learning Model for Medical Image Analysis. *International Journal of Information Technology*, 14(1), 95. 10.1007/S41870-021-00850-4
- Chittajallu, S. M., Mandalaneni, N. L. D., Parasa, D., & Bano, S. (2019). Classification Of Binary Fracture Using CNN. Paper Presented at the *2019 Global Conference for Advancement in Technology (GCAT)*, 1–5.
- Chład, P., & Ogiela, M. R. (2023). Deep Learning and Cloud-Based Computation for Cervical Spine Fracture Detection System. *Electronics*, 12(9), 2056.
- Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A Review of Medical Image Data Augmentation Techniques for Deep Learning Applications. *Wiley*. 10.1111/1754-9485.13261
- Chua, L. O. (1997). CNN: A Vision of Complexity. *International Journal of Bifurcation and Chaos*, 7(10), 2219–2425.

- Cui, R., & Liu, M. (2019). RNN-Based Longitudinal Analysis for Diagnosis of Alzheimer's Disease. *Computerized Medical Imaging and Graphics*, 73, 1. 10.1016/J.Compmedimag.2019.01.005
- Ćurković, S., Antabak, A., Halužan, D., Luetić, T., Prlić, I., & Šiško, J. (2015). Medical Thermography (Digital Infrared Thermal Imaging–DITI) In Paediatric Forearm Fractures–A Pilot Study. *Injury*, 46, S36–S39.
- D'Antonoli, T. A. (2020). Ethical Considerations for Artificial Intelligence: An Overview of the Current Radiology Landscape. *Diagnostic And Interventional Radiology*, 26(5), 504.
- Das, T., & Saha, G. (2020). Addressing Big Data Issues Using RNN Based Techniques. *Journal Of Information and Optimization Sciences*, 40(8), 1773. 10.1080/02522667.2019.1703268
- Diakides, M., Bronzino, J. D., & Peterson, D. R. (2012). Medical Infrared Imaging: Principles And Practices. *CRC Press*.
- Dimitriou, R., Tsiridis, E., & Giannoudis, P. V. (2005). Current Concepts of Molecular Aspects of Bone Healing. *Injury*, 36(12), 1392–1404.
- Ding, T., Shi, K., Pan, Z., & Ding, C. (2024). AI-Based Automated Breast Cancer Segmentation in Ultrasound Imaging Based on Attention Gated Multi Resu-Net. *Peerj Computer Science*, 10, E2226.
- Ding, Y., Chen, F., Zhao, Y., Wu, Z., Zhang, C., & Wu, D. (2019). A Stacked Multi-Connection Simple Reducing Net for Brain Tumor Segmentation. *IEEE Access*, 7, 104011. 10.1109/Access.2019.2926448
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). *AN IMAGE IS WORTH 16X16 WORDS: TRANSFORMERS FOR IMAGE RECOGNITION AT SCALE*.
- Dritsas, E., & Trigka, M. (2022). Machine Learning Techniques for Chronic Kidney Disease Risk Prediction. *Big Data and Cognitive Computing*, 6(3), 98.
- Du, M., Lin, Y., Chen, W., & Zhao, H. (2022). Advances In the Application of Ultrasound for Fracture Diagnosis and Treatment. *European Review for Medical & Pharmacological Sciences*, 26(21)
- Ehab, W., & Li, Y. (2023). Performance Analysis of Unet and Variants for Medical Image Segmentation. *Arxiv Preprint Arxiv:2309.13013*,

- Elhassan, T. A., Mohd Rahim, M. S., Siti Zaiton, M. H., Swee, T. T., Alhaj, T. A., Ali, A., & Aljurf, M. (2023). Classification Of Atypical White Blood Cells in Acute Myeloid Leukemia Using a Two-Stage Hybrid Model Based on Deep Convolutional Autoencoder and Deep Convolutional Neural Network. *MDPI AG*. 10.3390/Diagnostics13020196
- Epema, A. C., Spanjer, M. J., Ras, L., Kelder, J. C., & Sanders, M. (2019). Point-Of-Care Ultrasound Compared with Conventional Radiographic Evaluation in Children with Suspected Distal Forearm Fractures in the Netherlands: A Diagnostic Accuracy Study. *Emergency Medicine Journal*, 36(10), 613–616.
- Eschweiler, J., Li, J., Quack, V., Rath, B., Baroncini, A., Hildebrand, F., & Migliorini, F. (2022). Anatomy, Biomechanics, And Loads of The Wrist Joint. *Life*, 12(2), 188.
- Fane De Salis, A., Saatchi, R., & Dimitri, P. (2018). Evaluation Of High-Resolution Thermal Imaging to Determine the Effect of Vertebral Fractures on Associated Skin Surface Temperature in Children with Osteogenesis Imperfecta. *Medical & Biological Engineering & Computing*, 56(9), 1633–1643. 10.1007/S11517-018-1806-3
- Fang, Q., Jiang, A., Liu, M., & Zhao, S. (2025). Faster R-CNN Model for Target Recognition and Diagnosis of Scapular Fractures. *Journal Of Bone Oncology*, 51, 100664.
- Firoozabadi, R., Morshed, S., Engelke, K., Prevrhal, S., Fierlinger, A., Miclau III, T., & Genant, H. K. (2008). Qualitative And Quantitative Assessment of Bone Fragility and Fracture Healing Using Conventional Radiography and Advanced Imaging Technologies-Focus on Wrist Fracture. *Journal Of Orthopaedic Trauma*, 22, S83–S90.
- Fassold, H., Rosner, J., Schallauer, P., & Bailer, W. (2009). Realtime KLT feature point tracking for high-definition video. *GraVisMa*,
- Florencio-Silva, R., Sasso, G. R. D. S., Sasso-Cerri, E., Simões, M. J., & Cerri, P. S. (2015). Biology Of Bone Tissue: Structure, Function, And Factors That Influence Bone Cells. *Biomed Research International*, 2015(1), 421746.
- Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). GAN-Based Synthetic Medical Image Augmentation for Increased CNN Performance in Liver Lesion Classification. *Neurocomputing*, 321, 321–331.
- Fritz, A. V., Martin, A. K., Belli, E., Clendenen, S. R., & Clendenen, S. (2022). Intraoperative Epicardial Ultra-High Frequency Ultrasound in Coronary Artery Bypass Grafting Surgery. *Cureus*, 14(2)

- Galbusera, F., & Cina, A. (2024a). Image Annotation and Curation in Radiology: An Overview for Machine Learning Practitioners. *European Radiology Experimental*, 8(1), 11.
- Galbusera, F., & Cina, A. (2024b). Image Annotation and Curation in Radiology: An Overview for Machine Learning Practitioners. *European Radiology Experimental*, 8(1), 11.
- Gao, J., Zhao, W., Li, P., Huang, W., & Chen, Z. (2022). LEGAN: A Light and Effective Generative Adversarial Network for Medical Image Synthesis. *Computers In Biology and Medicine*, 14810.1016/J.Compbimed.2022.105878
- Gao, Y., Zeng, S., Xu, X., Li, H., Yao, S., Song, K., ... & Gao, Q. (2022). Deep learning-enabled pelvic ultrasound images for accurate diagnosis of ovarian cancer in China: a retrospective, multicentre, diagnostic study. *The Lancet Digital Health*, 4(3), e179-e187.
- Garbaz, A., Oukdach, Y., Charfi, S., El Ansari, M., Koutti, L., & Salihoun, M. (2024). MLFA-Net: A Multi-Level Feature Assembly Unet for Medical Image Segmentation. *Methods*, 232, 52. 10.1016/J.Ymeth.2024.10.010
- Garbin, C., Zhu, X., & Marques, O. (2020). Dropout Vs. Batch Normalization: An Empirical Study of Their Impact to Deep Learning. *Multimedia Tools and Applications*, 79(19), 12777–12815.
- Ghiasi, M. S., Chen, J., Vaziri, A., Rodriguez, E. K., & Nazarian, A. (2017). Bone Fracture Healing in Mechanobiological Modeling: A Review of Principles and Methods. *Bone Reports*, 6, 87–100.
- Gonzalez, R. C. (2009). *Digital Image Processing*. Pearson Education India.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2020). Generative Adversarial Networks. *Communications Of The ACM*, 63(11), 139–144.
- Guan, B., Yao, J., & Zhang, G. (2024). An Enhanced Vision Transformer with Scale-Aware and Spatial-Aware Attention for Thighbone Fracture Detection. *Neural Computing and Applications*, 36(19), 11425–11438.
- Guan, Q., Wang, Y., Du, J., Qin, Y., Lu, H., Xiang, J., & Wang, F. (2019). Deep Learning Based Classification of Ultrasound Images for Thyroid Nodules: A Large Scale of Pilot Study. *Annals Of Translational Medicine*, 7(7), 137.

- Gulshan, & Arora, A. S. (2024). Automated Prediction of Diabetes Mellitus Using Infrared Thermal Foot Images: Recurrent Neural Network Approach. *Biomedical Physics & Engineering Express*, 10(2)10.1088/2057-1976/Ad2479
- Guo, H., & Li, Z. (2019). Application Of Medical Infrared Thermal Imaging in the Diagnosis of Human Internal Focus. *Infrared Physics & Technology*, 101, 127–132.
- Gupta, S., & Sharma, D. (2024). Bone Fracture Classification Using Transfer Learning. *Arxiv Preprint Arxiv:2406.15958*,
- Haluzan, D., Davila, S., Antabak, A., Dobric, I., Stipic, J., Augustin, G., Ehrenfreund, T., & Prlic, I. (2015). Thermal Changes During Healing of Distal Radius Fractures—Preliminary Findings. *Injury*, 46, S103–S106.
- Hamdan, H., & Garibaldi, J. M. (2010). Adaptive Neuro-Fuzzy Inference System (ANFIS) In Modelling Breast Cancer Survival. Paper Presented at the *International Conference on Fuzzy Systems*, 1–8.
- Hamghalam, M., Wang, T., & Lei, B. (2020). High Tissue Contrast Image Synthesis Via Multistage Attention-GAN: Application to Segmenting Brain MR Scans. *Neural Networks*, 132, 43. 10.1016/J.Neunet.2020.08.014
- Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., & Tao, D. (2023). A Survey on Vision Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 87. 10.1109/Tpami.2022.3152247
- Hareendranathan, A. R., Tripathi, A., Panicker, M. R., Zhang, J., Boora, N., & Jaremko, J. (2023). Deep Learning Approach for Automatic Wrist Fracture Detection Using Ultrasound Bone Probability Maps. *SN Comprehensive Clinical Medicine*, 5(1), 276.
- Harrington, P. D. B. (1993). Sigmoid Transfer Functions in Backpropagation Neural Networks. *Analytical Chemistry*, 65(15), 2167–2168.
- Hauser, R. A., Dolan, E. E., Phillips, H. J., Newlin, A. C., Moore, R. E., & Woldin, B. A. (2013). Ligament Injury and Healing: A Review of Current Clinical Diagnostics and Therapeutics. *Open Rehabil J*, 6(1), 1–20.
- Helmbold, D. P., & Long, P. M. (2018). Surprising Properties of Dropout in Deep Networks. *Journal Of Machine Learning Research*, 18(200), 1–28.

- Hemanth, D. J., Vijila, C. K. S., & Anitha, J. (2011). Application Of Neuro-Fuzzy Model for Mr Brain Tumor Image Classification. *International Journal of Biomedical Soft Computing and Human Sciences: The Official Journal of the Biomedical Fuzzy Systems Association*, 16(1), 95–102.
- Henson, W. H., Li, X., Lin, Z., Guo, L., Mazzá, C., & Dall'ara, E. (2024). *Automatic Segmentation of Lower Limb Muscles from MR Images of Post-Menopausal Women Based on Deep Learning and Data Augmentation*. Public Library of Science (Plos). 10.1371/Journal.Pone.0299099
- Hernandez-Guedes, A., Santana-Perez, I., Arteaga-Marrero, N., Fabelo, H., Callico, G. M., & Ruiz-Alzola, J. (2022). Performance Evaluation of Deep Learning Models for Image Classification Over Small Datasets: Diabetic Foot Case Study. *IEEE Access*, 10, 124373–124386.
- Hollinger, J. O., & Guelcher, S. A. (2012). *An Introduction to Biomaterials*. CRC Press/Taylor & Francis Boca Raton, Florida. USA.
- Hossein Ghayoumi Zadeh, Javad Haddadnia, Maryam Hashemian, & Kazem Hassanpour. (2013). Diagnosis Of Breast Cancer Using a Combination of Genetic Algorithm and Artificial Neural Network in Medical Infrared Thermal Imaging. *Iranian Journal of Medical Physics*, 9(4), 265–274. 10.22038/Ijmp.2013.470
- Hosseini, M. S., & Zekri, M. (2012). Review Of Medical Image Classification Using the Adaptive Neuro-Fuzzy Inference System. *Journal Of Medical Signals & Sensors*, 2(1), 49–60.
- Huang, K., Yang, Y., Huang, Z., Liu, Y., & Lee, S. (2023). Retinal Vascular Image Segmentation Using Improved Unet Based on Residual Module. *MDPI AG*. 10.3390/Bioengineering10060722
- Iglesias, G., Talavera, E., González-Prieto, Á, Mozo, A., & Gómez-Canaval, S. (2023). Data Augmentation Techniques in Time Series Domain: A Survey and Taxonomy. *Neural Computing and Applications*, 35(14), 10123–10145.
- Iqbal, S., Khan, T. M., Naqvi, S. S., Naveed, A., & Meijering, E. (2025). Tbconvl-Net: A Hybrid Deep Learning Architecture for Robust Medical Image Segmentation. *Pattern Recognition*, 15810.1016/J.Patcog.2024.111028
- Iwana, B. K., & Uchida, S. (2021). An Empirical Survey of Data Augmentation for Time Series Classification with Neural Networks. *Plos One*, 16(7), E0254841.

- Jahagirdar, R., & Scammell, B. E. (2009). Principles Of Fracture Healing and Disorders of Bone Union. *Surgery (Oxford)*, 27(2), 63–69.
- Jang, J. (1993). ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Transactions on Systems, Man, And Cybernetics*, 23(3), 665–685.
- Järvinen, T. A., Järvinen, T. L., Kääriäinen, M., Kalimo, H., & Järvinen, M. (2005). Muscle Injuries: Biology And Treatment. *The American Journal of Sports Medicine*, 33(5), 745–764.
- Jayaraj, R., & Lokesh, S. (2023). Automatic Image Annotation Using Adaptive Convolutional Deep Learning Model. *Intelligent Automation & Soft Computing*, 36(1)
- Jennison, T., & Brinsden, M. (2019). Fracture Admission Trends in England Over a Ten-Year Period. *The Annals of the Royal College of Surgeons of England*, 101(3), 208–214.
- Jeong, J. J., Tariq, A., Adejumo, T., Trivedi, H., Gichoya, J. W., & Banerjee, I. (2022). Systematic Review of Generative Adversarial Networks (Gans) For Medical Image Classification and Segmentation. *Journal Of Digital Imaging*, 35(2), 137. 10.1007/S10278-021-00556-W
- Jiang, L., Xu, C., Bai, Y., Liu, A., Gong, Y., Wang, Y., & Deng, H. (2024). Autosurv: Interpretable Deep Learning Framework for Cancer Survival Analysis Incorporating Clinical and Multi-Omics Data. *NPJ Precision Oncology*, 8(1), 4.
- Jothi, R., & Jayanthi, K. (2024). A Deep Multi-Filtering Network Based on Dilated Mask RCNN Framework with Self-Attention RNN Model for Segmenting and Classifying Bone Fracture. *Multimedia Tools and Applications*, 1–27.
- Juneja, M., Kaur Saini, S., Kaul, S., Acharjee, R., Thakur, N., & Jindal, P. (2021). Denoising Of Magnetic Resonance Imaging Using Bayes Shrinkage Based Fused Wavelet Transform and Autoencoder Based Deep Learning Approach. *Biomedical Signal Processing and Control*, 6910.1016/J.Bspc.2021.102844
- Kalogiouri, N. P., & Samanidou, V. F. (2021). A Validated Ultrasound-Assisted Extraction Coupled with Spe-Hplc-Dad for the Determination of Flavonoids in By-Products of Plant Origin: An Application Study for the Valorization of the Walnut Septum Membrane. *Molecules*, 26(21), 6418.
- Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2021). Enhancing Diagnostic Accuracy in Medical Imaging Through Convolutional Neural Networks and Transfer Learning Algorithms. *International Journal of AI And ML*, 2(3)

- Kani, J. N., & Elsheikh, A. H. (2017). DR-RNN: A Deep Residual Recurrent Neural Network for Model Reduction. *Arxiv Preprint Arxiv:1709.00939*,
- Kar, S., Das, S., & Ghosh, P. K. (2014). Applications Of Neuro Fuzzy Systems: A Brief Review and Future Outline. *Applied Soft Computing, 15*, 243–259.
- Karaboga, D., & Kaya, E. (2018). *Adaptive Network Based Fuzzy Inference System (ANFIS) Training Approaches: A Comprehensive Survey*. Springer Science and Business Media LLC. 10.1007/S10462-017-9610-2
- Kassem, M. A., Naguib, S. M., Hamza, H. M., Fouda, M. M., Saleh, M. K., & Hosny, K. M. (2023). Explainable Transfer Learning-Based Deep Learning Model for Pelvis Fracture Detection. *International Journal of Intelligent Systems, 2023*(1), 3281998.
- Kastberger, G., & Stachl, R. (2003). Infrared Imaging Technology and Biological Applications. *Behavior Research Methods, Instruments, & Computers, 35*(3), 429–439. 10.3758/BF03195520
- Kellerova, E., Delius, W., Olerud, S., & Ström, G. (1970). Changes In the Muscle and Skin Blood Flow Following Lowfr Leg Fracture in Man. *Acta Orthopaedica Scandinavica, 41*(3), 249–260.
- Keszyüs, D., Brucher, S., & Keszyüs, T. (2022). Use Of Infrared Thermography in Medical Diagnostics: A Scoping Review Protocol. *BMJ Open, 12*(4), E059833.
- Khan, M. I., Amin, J., Shehzad, M. A., & Iqbal, S. (2025). AN INNOVATIVE HYBRID MODEL FOR ELBOW BONE FRACTURE DETECTION: INTEGRATING VIT AND CNN. *Contemporary Journal of Social Science Review, 3*(1), 1–21.
- Khan, R. A., Luo, Y., & Wu, F. (2022). RMS-Unet: Residual Multi-Scale Unet for Liver and Lesion Segmentation. *Artificial Intelligence in Medicine, 124*10.1016/J.Artmed.2021.102231
- Khosla, S., Melton III, L. J., Dekutoski, M. B., Achenbach, S. J., Oberg, A. L., & Riggs, B. L. (2003). Incidence Of Childhood Distal Forearm Fractures Over 30 Years: A Population-Based Study. *Jama, 290*(11), 1479–1485.
- Kim, D., Park, E., Ku, K., Hwang, S. J., Hwang, K. T., Lee, C., & Yoon, G. H. (2023). Application Of Stacked Autoencoder for Identification of Bone Fracture. *Journal of the Mechanical Behavior of Biomedical Materials, 146*, 106077.
- Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). *Transfer Learning for Medical Image Classification: A Literature Review*. Springer Science and Business Media LLC. 10.1186/S12880-022-00793-7

- Kim, J., Hong, J., & Park, H. (2018). Prospects Of Deep Learning for Medical Imaging. *Precision And Future Medicine*, 2(2), 37. 10.23838/Pfm.2018.00030
- Kim, M., Zuallaert, J., & De Neve, W. (2017). Few-Shot Learning Using a Small-Sized Dataset of High-Resolution Fundus Images for Glaucoma Diagnosis. Paper Presented at the *Proceedings of the 2nd International Workshop on Multimedia for Personal Health and Health Care*, 89–92.
- Kokol, P., Kokol, M., & Zagoranski, S. (2022). Machine Learning on Small Size Samples: A Synthetic Knowledge Synthesis. *Science Progress*, 105(1), 00368504211029777.
- Kong, Z., Zhang, M., Zhu, W., Yi, Y., Wang, T., & Zhang, B. (2022). Data Enhancement Based on M2-Unet for Liver Segmentation in Computed Tomography. *Biomedical Signal Processing and Control*, 7910.1016/J.Bspc.2022.104032
- Kora, P., Ooi, C. P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W. Y., Meenakshi, K., Swaraja, K., Plawiak, P., & Rajendra Acharya, U. (2021). Transfer Learning Techniques for Medical Image Analysis: A Review. *Biocybernetics And Biomedical Engineering*, 42(1), 79. 10.1016/J.Bbe.2021.11.004
- Krithika Alias Anbudevi, M., & Suganthi, K. (2022). Review Of Semantic Segmentation of Medical Images Using Modified Architectures Of UNET. *Diagnostics*, 12(12)10.3390/Diagnostics12123064
- Kumar, R., Al-Turjman, F., Srinivas, L., Braveen, M., & Ramakrishnan, J. (2023a). ANFIS For Prediction of Epidemic Peak and Infected Cases For COVID-19 In India. *Neural Computing and Applications*, 35(10), 7207–7220.
- Kumar, R., Al-Turjman, F., Srinivas, L., Braveen, M., & Ramakrishnan, J. (2023b). ANFIS For Prediction of Epidemic Peak and Infected Cases For COVID-19 In India. *Neural Computing and Applications*, 35(10), 7207–7220.
- Kumar, S., Choudhary, S., Jain, A., Singh, K., Ahmadian, A., & Bajuri, M. Y. (2023). Brain Tumor Classification Using Deep Neural Network and Transfer Learning. *Brain Topography*, 36(3), 305. 10.1007/S10548-023-00953-0
- Kumar, T., Brennan, R., Mileo, A., & Bendeche, M. (2024). Image Data Augmentation Approaches: A Comprehensive Survey and Future Directions. *IEEE Access*,

- Kumarganesh, S., & Suganthi, M. (2018). An Enhanced Medical Diagnosis Sustainable System for Brain Tumor Detection and Segmentation Using ANFIS Classifier. *Current Medical Imaging*, 14(2), 271–279.
- Kumarganesh, S., & Suganthi, M. (2023). *An Enhanced Medical Diagnosis Sustainable System for Brain Tumor Detection and Segmentation Using ANFIS Classifier*. Bentham Science Publishers Ltd. 10.2174/1573405613666161216122938
- Lahiri, B. B., Bagavathiappan, S., Jayakumar, T., & Philip, J. (2012). Medical Applications of Infrared Thermography: A Review. *Infrared Physics & Technology*, 55(4), 221–235.
- Lalkhen, A. G., & McCluskey, A. (2008). Clinical Tests: Sensitivity And Specificity. *Continuing Education in Anaesthesia, Critical Care & Pain*, 8(6), 221–223.
- Lastayo, P. C., Winters, K. M., & Hardy, M. (2003). Fracture Healing: Bone Healing, Fracture Management, And Current Concepts Related to the Hand. *Journal Of Hand Therapy*, 16(2), 81–93.
- Lewis, J. P. (1995). Fast Template Matching. Paper Presented at the *Vision Interface*, , 95(120123) 15–19.
- Li, B., Hou, Y., & Che, W. (2022). Data Augmentation Approaches in Natural Language Processing: A Survey. *Ai Open*, 3, 71–90.
- Li, D., Fu, Z., & Xu, J. (2020). Stacked-Autoencoder-Based Model For COVID-19 Diagnosis on CT Images. *Applied Intelligence*, 51(5), 2805. 10.1007/S10489-020-02002-W
- Li, J., Wang, D., Li, S., Zhang, M., Song, C., & Chen, X. (2019). Deep Learning Based Adaptive Sequential Data Augmentation Technique for the Optical Network Traffic Synthesis. *Optics Express*, 27(13), 18831–18847.
- Li, Y., Zhang, K., Shi, W., Miao, Y., & Jiang, Z. (2021). A Novel Medical Image Denoising Method Based on Conditional Generative Adversarial Network. *Computational And Mathematical Methods in Medicine*, 2021, 1. 10.1155/2021/9974017
- Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A Survey of Convolutional Neural Networks: Analysis, Applications, And Prospects. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12), 6999–7019.
- Liao, Z., Liao, K., Shen, H., Van Boxel, M. F., Prijs, J., Jaarsma, R. L., Doornberg, J. N., Van Den Hengel, A., & Verjans, J. W. (2022). CNN Attention Guidance for Improved Orthopedics

- Radiographic Fracture Classification. *IEEE Journal of Biomedical and Health Informatics*, 26(7), 3139–3150.
- Lin, Z., Dall'ara, E., & Guo, L. (2024). A Novel Mean Shape Based Post-Processing Method for Enhancing Deep Learning Lower-Limb Muscle Segmentation Accuracy. *Public Library of Science (Plos)*. 10.1371/Journal.Pone.0308664
- Lin, Z., Henson, W. H., Dowling, L., Walsh, J., Dall'ara, E., & Guo, L. (2024). Automatic Segmentation of Skeletal Muscles from MR Images Using Modified U-Net and a Novel Data Augmentation Approach. *Frontiers Media SA*. 10.3389/Fbioe.2024.1355735
- Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable Ai: A Review of Machine Learning Interpretability Methods. *Entropy*, 23(1), 18.
- Lindau, T. (2012). A New Definition of Wrist Sprain Necessary After Findings in a Prospective MRI Study.
- Little, J. T., Kliensky, N. B., Chaturvedi, A., Soral, A., & Chaturvedi, A. (2014). Pediatric Distal Forearm and Wrist Injury: An Imaging Review. *Radiographics*, 34(2), 472–490.
- Liu, X., Song, L., Liu, S., & Zhang, Y. (2021). A Review of Deep-Learning-Based Medical Image Segmentation Methods. *Sustainability*, 13(3), 1224.
- Liu, X., Karagoz, G., & Meratnia, N. (2024). Analyzing The Impact of Data Augmentation on the Explainability of Deep Learning-Based Medical Image Classification. *Machine Learning and Knowledge Extraction*, 7(1), 1.
- Loi, F., Córdova, L. A., Pajarinen, J., Lin, T., Yao, Z., & Goodman, S. B. (2016). Inflammation, Fracture and Bone Repair. *Bone*, 86, 119–130.
- Ma, J., Hu, C., Zhou, P., Jin, F., Wang, X., & Huang, H. (2023). Review Of Image Augmentation Used in Deep Learning-Based Material Microscopic Image Segmentation. *Applied Sciences*, 13(11), 6478.
- Mahapatra, D., Bozorgtabar, B., & Garnavi, R. (2019). Image Super-Resolution Using Progressive Generative Adversarial Networks for Medical Image Analysis. *Computerized Medical Imaging and Graphics*, 71, 30. 10.1016/J.Compmedimag.2018.10.005
- Maharana, K., Mondal, S., & Nemade, B. (2022). A Review: Data Pre-Processing and Data Augmentation Techniques. *Global Transitions Proceedings*, 3(1), 91–99.

- Mamat, N., Othman, M. F., Abdulghafor, R., Alwan, A. A., & Gulzar, Y. (2023). Enhancing Image Annotation Technique of Fruit Classification Using a Deep Learning Approach. *Sustainability*, *15*(2), 901.
- Mamoowala, N., Johnson, N. A., & Dias, J. J. (2019). Trends In Paediatric Distal Radius Fractures: An Eight-Year Review from a Large UK Trauma Unit. *The Annals of the Royal College of Surgeons of England*, *101*(4), 297–303.
- Marcinkevičs, R., & Vogt, J. E. (2023). Interpretable And Explainable Machine Learning: A Methods-Centric Overview with Concrete Examples. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *13*(3), E1493.
- Marenzana, M., & Arnett, T. R. (2013). The Key Role of the Blood Supply to Bone. *Bone Research*, *1*(1), 203–215.
- Marsell, R., & Einhorn, T. A. (2011). The Biology of Fracture Healing. *Injury*, *42*(6), 551–555.
- Masoudi, A., & Naraghi, L. (2022). Point-Of-Care Ultrasound for Diagnosis and Pain Control of Sternal Fracture. *Cureus*, *14*(3)
- Masumoto, R., Eguchi, Y., Takeuchi, H., Inage, K., Narita, M., Shiga, Y., Inoue, M., Toshi, N., Tokeshi, S., Okuyama, K., Ohyama, S., Suzuki, N., Maki, S., Furuya, T., Ohtori, S., & Orita, S. (2024). Automatic Generation of Diffusion Tensor Imaging for the Lumbar Nerve Using Convolutional Neural Networks. *Magnetic Resonance Imaging*, *114*10.1016/J. Mri.2024.110237
- Mayeta-Revilla, L., Cavieres, E. P., Salinas, M., Mellado, D., Ponce, S., Torres Moyano, F., Chabert, S., Querales, M., Sotelo, J., & Salas, R. (2025). Radiomics-Driven Neuro-Fuzzy Framework for Rule Generation to Enhance Explainability In MRI-Based Brain Tumor Segmentation. *Frontiers In Neuroinformatics*, *19*, 1550432.
- Meor Yahaya, M. S., & Teo, J. (2023). Data Augmentation Using Generative Adversarial Networks for Images and Biomarkers in Medicine and Neuroscience. *Frontiers In Applied Mathematics and Statistics*, *9*, 1162760.
- Morid, M. A., Borjali, A., & Del Fiol, G. (2021). A Scoping Review of Transfer Learning Research on Medical Image Analysis Using Imagenet. *Computers In Biology and Medicine*, *128*, 104115.
- Mumuni, A., & Mumuni, F. (2022). Data Augmentation: A Comprehensive Survey of Modern Approaches. *Array*, *16*, 100258.

- Munsayac, F. E. T., Alonzo, L. M. B., Lindo, D. E. G., Baldovino, R. G., & Bugtai, N. T. (2017). Implementation Of a Normalized Cross-Correlation Coefficient-Based Template Matching Algorithm in Number System Conversion. Paper Presented at the *2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 1–4.
- Murphy, P. J., Myers, B. L., & Badia, P. (1996). Nonsteroidal Anti-Inflammatory Drugs Alter Body Temperature and Suppress Melatonin in Humans. *Physiology & Behavior*, *59*(1), 133–139.
- Mutasa, S., Varada, S., Goel, A., Wong, T. T., & Rasiej, M. J. (2020). Advanced Deep Learning Techniques Applied to Automated Femoral Neck Fracture Detection and Classification. *Journal Of Digital Imaging*, *33*(5), 1209–1217.
- Mutepfe, F., Kalejahi, B. K., Meshgini, S., & Danishvar, S. (2021). Generative Adversarial Network Image Synthesis Method for Skin Lesion Generation and Classification. *Journal Of Medical Signals & Sensors*, *11*(4), 237. 10.4103/Jmss.Jmss_53_20
- Nahangi, H., & Chaparian, A. (2015). Assessment Of Radiation Risk to Pediatric Patients Undergoing Conventional X-Ray Examinations. *Radioprotection*, *50*(1), 19. 10.1051/Radiopro/2014023
- Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. Paper Presented at the *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 807–814.
- Nanni, L., Paci, M., Brahnam, S., & Lumini, A. (2021). Comparison Of Different Image Data Augmentation Approaches. *Journal Of Imaging*, *7*(12), 254.
- Nasser, Y., El Hassouni, M., Brahim, A., Toumi, H., Lespessailles, E., & Jennane, R. (2017). Diagnosis Of Osteoporosis Disease from Bone X-Ray Images with Stacked Sparse Autoencoder and SVM Classifier. Paper Presented at the *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 1–5.
- Natrajan, P., Bhat, R. R., Remadevi, R., Joseph, I. R., Vijayalakshmi, S., & Paulose, T. D. (2021). Comparative Study to Evaluate the Effect of Ultrasound-Guided Pericapsular Nerve Group Block Versus Fascia Iliaca Compartment Block on the Postoperative Analgesic Effect in Patients Undergoing Surgeries for Hip Fracture Under Spinal Anesthesia. *Anesthesia Essays and Research*, *15*(3), 285–289.

- Nguyen, T. P., Chae, D., Choi, S. H., Jeong, K., & Yoon, J. (2023). Enhancement Of Hip X-Ray with Convolutional Autoencoder for Increasing Prediction Accuracy of Bone Mineral Density. *Bioengineering*, *10*(10), 1169.
- Nyary, T., & Scammell, B. E. (2018). Principles Of Bone and Joint Injuries and Their Healing. *Surgery (Oxford)*, *36*(1), 7–14.
- O’ Sullivan, E., Van De Lande, L. S., Papaioannou, A., Breakey, R. W. F., Jeelani, N. O., Ponniah, A., Duncan, C., Schievano, S., Khonsari, R. H., Zafeiriou, S., & Dunaway, D. J. (2022). Convolutional Mesh Autoencoders for the 3-Dimensional Identification Of FGFR-Related Craniosynostosis. *Scientific Reports*, *12*(1)10.1038/S41598-021-02411-Y
- Obayya, M., Saeed, M. K., Alruwais, N., Alotaibi, S. S., Assiri, M., & Salama, A. S. (2023). Hybrid Metaheuristics with Deep Learning-Based Fusion Model for Biomedical Image Analysis. *IEEE Access*, *11*, 117149. 10.1109/Access.2023.3326369
- Obuchowski, N. A., & Bullen, J. A. (2018). Receiver Operating Characteristic (ROC) Curves: Review of Methods with Applications in Diagnostic Medicine. *Physics In Medicine & Biology*, *63*(7), 07TR01.
- Odusami, M., Maskeliūnas, R., Damaševičius, R., & Krilavičius, T. (2021). Analysis Of Features of Alzheimer’s Disease: Detection of Early Stage from Functional Brain Changes in Magnetic Resonance Images Using a Finetuned Resnet18 Network. *Diagnostics*, *11*(6)10.3390/Diagnostics11061071
- Ogunleye, B., Sharma, H., & Shobayo, O. (2024). Sentiment Informed Sentence BERT-Ensemble Algorithm for Depression Detection. *Big Data and Cognitive Computing*, *8*(9)10.3390/Bdcc8090112
- Oryan, A., Alidadi, S., & Moshiri, A. (2013). Current Concerns Regarding Healing of Bone Defects. *Hard Tissue*, *2*(2)10.13172/2050-2303-2-2-374
- Oryan, A., Monazzah, S., & Bigam-Sadegh, A. (2015). Bone Injury and Fracture Healing Biology. *Biomedical And Environmental Sciences*, *28*(1), 57–71. 10.3967/Bes2015.006
- Owen, R., Ramlakhan, S., Saatchi, R., & Burke, D. (2017). Development Of a High-Resolution Infrared Thermographic Imaging Method as a Diagnostic Tool for Acute Undifferentiated Limp in Young Children. *Medical & Biological Engineering & Computing*, *56*(6), 1115–1125. 10.1007/S11517-017-1749-0

- Owen, R., & Ramlakhan, S. (2017). Infrared Thermography in Paediatrics: A Narrative Review of Clinical Use. *BMJ Paediatrics Open*, 1(1), E000080.
- Oza, P., Sharma, P., Patel, S., Adedoyin, F., & Bruno, A. (2022). Image Augmentation Techniques for Mammogram Analysis. *Journal Of Imaging*, 8(5), 141.
- Pasha, A., Ahmed, S. T., Painam, R. K., Mathivanan, S. K., Mallik, S., & Qin, H. (2024). Leveraging ANFIS With Adam and PSO Optimizers for Parkinson's Disease. *Elsevier BV*. 10.1016/J.Heliyon.2024.E30241
- Pasini, A. (2015). Artificial Neural Networks for Small Dataset Analysis. *Journal Of Thoracic Disease*, 7(5), 953.
- Pham, T., Luong, C., Visani, M., & Hoang, V. (2018). Deep CNN And Data Augmentation for Skin Lesion Classification. Paper Presented at the *Asian Conference on Intelligent Information and Database Systems*, 573–582.
- Pietrolaj, M., & Blok, M. (2024). Resource Constrained Neural Network Training. *Springer Science and Business Media LLC*. 10.1038/S41598-024-52356-1
- Piffer, S., Ubaldi, L., Tangaro, S., Retico, A., & Talamonti, C. (2024). Tackling The Small Data Problem in Medical Image Classification with Artificial Intelligence: A Systematic Review. *Progress In Biomedical Engineering*, 6(3), 032001.
- Pilehvari, S., Morgan, Y., & Peng, W. (2024). An Analytical Review on The Use of Artificial Intelligence and Machine Learning in Diagnosis, Prediction, And Risk Factor Analysis of Multiple Sclerosis. *Multiple Sclerosis and Related Disorders*, 89, 105761. 10.1016/J.Msard.2024.105761
- Pilitsis, J. G., Lucas, D. R., & Rengachary, S. R. (2002). Bone Healing and Spinal Fusion. *Neurosurgical Focus*, 13(6), 1–6.
- Piotrowski, A. P., Napiorkowski, J. J., & Piotrowska, A. E. (2020). Impact Of Deep Learning-Based Dropout on Shallow Neural Networks Applied to Stream Temperature Modelling. *Earth-Science Reviews*, 201, 103076.
- Plastropoulos, A., Adiuku, N., Avdelidis, N. P., Ibarra-Castanedo, C., Maldague, X. P., & Servais, P. (2023). A Short Review on Infrared Thermography: History, Philosophy, Approaches, Standards, And Application Examples. *Thermosense: Thermal Infrared Applications XLV*, 12536, 5–15.

- Popescu, M., Balas, V. E., Perescu-Popescu, L., & Mastorakis, N. (2009). Multilayer Perceptron and Neural Networks. *WSEAS Transactions on Circuits and Systems*, 8(7), 579–588.
- Preston, D. L., Ron, E., Tokuoka, S., Funamoto, S., Nishi, N., Soda, M., Mabuchi, K., & Kodama, K. (2007). Solid Cancer Incidence in Atomic Bomb Survivors: 1958–1998. *Radiation Research*, 168(1), 1–64. 10.1667/RR0763.1
- Pu, Q., Xi, Z., Yin, S., Zhao, Z., & Zhao, L. (2024). Advantages Of Transformer and Its Application for Medical Image Segmentation: A Survey. *Springer Science and Business Media LLC*. 10.1186/S12938-024-01212-4
- Purwono, P., Wulandari, A. N. E., & Nisa, K. (2025). Explainable Artificial Intelligence (XAI) In Medical Imaging: Techniques, Applications, Challenges, And Future Directions. *Advanced Mechanical and Mechatronic Systems*, 1(1), 52–66.
- Qian, N. (1999). On The Momentum Term in Gradient Descent Learning Algorithms. *Neural Networks*, 12(1), 145–151.
- Rahman, H., Bukht, T. F. N., Imran, A., Tariq, J., Tu, S., & Alzahrani, A. (2022). A Deep Learning Approach for Liver and Tumor Segmentation in CT Images Using Resunet. *Bioengineering*, 9(8)10.3390/Bioengineering9080368
- Rajeev, R., Samath, J. A., & Karthikeyan, N. K. (2019). An Intelligent Recurrent Neural Network with Long Short-Term Memory (LSTM) BASED Batch Normalization for Medical Image Denoising. *Journal Of Medical Systems*, 43(8)10.1007/S10916-019-1371-9
- Ramamurthy, M., Krishnamurthi, I., Vimal, S., & Robinson, Y. H. (2020). Deep Learning Based Genome Analysis And NGS-RNA LL Identification with a Novel Hybrid Model. *Biosystems*, 19710.1016/J.Biosystems.2020.104211
- Ramlakhan, S. L., Saatchi, R., Sabir, L., Ventour, D., Shobayo, O., Hughes, R., & Singh, Y. (2022). Building Artificial Intelligence and Machine Learning Models: A Primer for Emergency Physicians. *Emergency Medicine Journal: EMJ*, 010.1136/Emermed-2022-212379
- Ramlakhan, S., Saatchi, R., Sabir, L., Singh, Y., Hughes, R., Shobayo, O., & Ventour, D. (2022). Understanding And Interpreting Artificial Intelligence, Machine Learning and Deep Learning in Emergency Medicine. *Emergency Medicine Journal: EMJ*, 010.1136/Emermed-2021-212068
- Rashid, T., Zia, M. S., Meraj, T., Rauf, H. T., & Kadry, S. (2023). A Minority Class Balanced Approach Using The DCNN-LSTM Method To Detect Human Wrist Fracture. *Life*, 13(1), 133.

- Reed, C., Saatchi, R., Burke, D., & Ramlakhan, S. (2020). Infrared Thermal Imaging as a Screening Tool for Paediatric Wrist Fractures. *Medical & Biological Engineering & Computing*, 58(7), 1549–1563. 10.1007/S11517-020-02167-Z
- Relyea-Chew, A., & Chew, F. S. (2019). Multiple Open Wrist Fractures and Dislocation of the Distal Radioulnar Joint from a Dog Bite Injury. *Radiology Case Reports*, 14(7), 837–841.
- Rennie, L., Court-Brown, C. M., Mok, J. Y., & Beattie, T. F. (2007). The Epidemiology of Fractures in Children. *Injury*, 38(8), 913–922.
- Richard, A. B., Friska, J., & Narayanan, K. L. (2025). Implementation Of ANFIS Assisted Modified CNN Classifier for Autism Spectrum Disorder Detection. Paper Presented at the 2025 *International Conference on Visual Analytics and Data Visualization (ICVADV)*, 1322–1326.
- Ring, E., & Ammer, K. (2012). Infrared Thermal Imaging in Medicine. *Physiological Measurement*, 33(3), R33.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lecture Notes in Computer Science*, 234. 10.1007/978-3-319-24574-4_28
- Roy, S. S. (2005). Design Of Adaptive Neuro-Fuzzy Inference System for Predicting Surface Roughness in Turning Operation. *Journal Of Scientific and Industrial Research*, 64(9), 653.
- Saatchi, R. (2024). Fuzzy Logic Concepts, Developments and Implementation. *MDPI AG*. 10.3390/Info15100656
- Saatchi, R., & Ramlakhan, S. (2023). Infrared Thermal Imaging Analysis in Screening for Toddler's Fracture: A Proof-Of-Concept Study. *Applied Sciences*, 13(24), 13299.
- Safdar, N. M., Banja, J. D., & Meltzer, C. C. (2020). Ethical Considerations in Artificial Intelligence. *European Journal of Radiology*, 122, 108768.
- Safonova, A., Ghazaryan, G., Stiller, S., Main-Knorn, M., Nendel, C., & Ryo, M. (2023). Ten Deep Learning Techniques to Address Small Data Problems with Remote Sensing. *International Journal of Applied Earth Observation and Geoinformation*, 125, 103569.
- Saied, M., Raafat, M., Yehia, S., & Khalil, M. M. (2023). Efficient Pulmonary Nodules Classification Using Radiomics and Different Artificial Intelligence Strategies. *Insights Into Imaging*, 14(1)10.1186/S13244-023-01441-6

- Salehin, I., & Kang, D. (2023). A Review on Dropout Regularization Approaches for Deep Neural Networks Within the Scholarly Domain. *Electronics*, 12(14), 3106.
- Sanchis-Sánchez, E., Salvador-Palmer, R., Codoñer-Franch, P., Martín, J., Vergara-Hernández, C., Blasco, J., Ballester, E., Sanchis, E., González-Peña, R., & Cibrián, R. (2015). Infrared Thermography Is Useful for Ruling Out Fractures in Paediatric Emergencies. *European Journal of Pediatrics*, 174(4), 493–499.
- Sandfort, V., Yan, K., Pickhardt, P. J., & Summers, R. M. (2019b). Data Augmentation Using Generative Adversarial Networks (CycleGAN) To Improve Generalizability in CT Segmentation Tasks. *Scientific Reports*, 9(1), 16884.
- Santana, M. A. D., Pereira, J. M. S., Silva, F. L. D., Lima, N. M. D., Sousa, F. N. D., Arruda, G. M. S. D., Lima, R. D. C. F. D., Silva, W. W. A. D., & Santos, W. P. D. (2018). Breast Cancer Diagnosis Based on Mammary Thermography and Extreme Learning Machines. *Research On Biomedical Engineering*, 34(1), 45–53. 10.1590/2446-4740.05217
- Sarmadi, A., Razavi, Z. S., Van Wijnen, A. J., & Soltani, M. (2024). Comparative Analysis of Vision Transformers and Convolutional Neural Networks in Osteoporosis Detection from X-Ray Images. *Scientific Reports*, 14(1)10.1038/S41598-024-69119-7
- Seens, H., Modarresi, S., Macdermid, J. C., Walton, D. M., & Grewal, R. (2021). Prevalence Of Bone Fractures Among Children and Adolescents with Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-Analysis. *BMC Pediatrics*, 21(1)10.1186/S12887-021-02821-X
- Sendher, R., & Ladd, A. L. (2013). The Scaphoid. *Orthopedic Clinics*, 44(1), 107–120.
- Shaikhina, T., & Khovanova, N. A. (2017). Handling Limited Datasets with Neural Networks in Medical Applications: A Small-Data Approach. *Artificial Intelligence in Medicine*, 75, 51–63.
- Sharkawy, A. (2020). Principle Of Neural Network and Its Main Types. *Journal Of Advances in Applied & Computational Mathematics*, 7(1), 8–19.
- Sharma, M., & Mukharjee, S. (2012a). *Brain Tumor Segmentation Using Hybrid Genetic Algorithm and Artificial Neural Network Fuzzy Inference System (ANFIS)*. Academy And Industry Research Collaboration Center (AIRCC). 10.5121/Ijfls.2012.2403
- Sharma, M., & Mukharjee, S. (2012b). Brain Tumor Segmentation Using Hybrid Genetic Algorithm and Artificial Neural Network Fuzzy Inference System (Anfis). *International Journal of Fuzzy Logic Systems*, 2(4), 31–42.

- Shatsky, I. G. (2017). Medical Exposure Risk Assessment For Radiographic Examinations of Children. *Radiacionnaâ Gigiena*, 10(2), 31–42. 10.21514/1998-426X-2017-10-2-31-42
- Shinohara, H., Shimpo, M., Watanabe, H., Toriumi, S., Komori, T., Hoshide, S., & Kario, K. (2021). Intravascular Ultrasound-Validated Mechanical Stress of the Aorta on Anomalous Origin of Coronary Artery—A Possible Sign of Angina—. *Circulation Journal*, 85(11), 2120.
- Shobayo, O., & Saatchi, R. (2025). Developments In Deep Learning Artificial Neural Network Techniques for Medical Image Analysis and Interpretation. *Diagnostics*, 15(9), 1072.
- Shobayo, O., Saatchi, R., & Ramlakhan, S. (2022). Infrared Thermal Imaging and Artificial Neural Networks to Screen for Wrist Fractures in Pediatrics. *Technologies*, 10(19)
- Shobayo, O., Saatchi, R., & Ramlakhan, S. (2024). Convolutional Neural Network to Classify Infrared Thermal Images of Fractured Wrists in Pediatrics. *Healthcare*, 12(10)10.3390/Healthcare12100994.
- Shobayo, O., Saatchi, R., & Ramlakhan, S. (2025). Adaptive Neuro-Fuzzy Inference System Framework for Paediatric Wrist Injury Classification. *Multimodal Technologies and Interaction*, 9(10), 104.
- Shorten, C., & Khoshgoftaar, T. M. (2019). A Survey on Image Data Augmentation for Deep Learning. *Journal Of Big Data*, 6(1), 1–48.
- Shu, X. -, Jin, F., Linet, M. S., Zheng, W., Clemens, J., Mills, J., & Gao, Y. -. (1994). Diagnostic X-Ray and Ultrasound Exposure and Risk of Childhood Cancer. *British Journal of Cancer*, 70(3), 531–536. 10.1038/Bjc.1994.340
- Shvetsova, N., Bakker, B., Fedulova, I., Schulz, H., & Dylov, D. V. (2021). Anomaly Detection in Medical Imaging with Deep Perceptual Autoencoders. *IEEE Access*, 9, 118571. 10.1109/Access.2021.3107163
- Sindhura D. N., Radhika M. Pai, Shyamasunder N. Bhat & Manohara Pai M. M. (2024). Deep Learning-Based Automated Spine Fracture Type Identification with Clinically Validated GAN Generated CT Images. *Cogent Engineering*, 11(1), 2295645, DOI: 10.1080/23311916.2023.2295645
- Singh, N. K., & Raza, K. (2021). Medical Image Generation Using Generative Adversarial Networks: A Review. (Pp. 77). Springer Singapore. 10.1007/978-981-15-9735-0_5

- Skansi, S. (2018). Autoencoders. *Introduction To Deep Learning: From Logical Calculus to Artificial Intelligence* (Pp. 153–163). Springer.
- Slaar, A., Bentohami, A., Kessels, J., Bijlsma, T. S., Van Dijkman, B. A., Maas, M., Wilde, J. C., Goslings, J. C., & Schep, N. W. (2012). The Role of Plain Radiography in Paediatric Wrist Trauma. *Insights Into Imaging*, 3(5), 513–517.
- Snehalatha, U., & Sangamithirai, K. (2021). Computer Aided Diagnosis of Obesity Based on Thermal Imaging Using Various Convolutional Neural Networks. *Biomedical Signal Processing and Control*, 63, 102233.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The Journal of Machine Learning Research*, 15(1), 1929–1958.
- Sriwong, K., Kerdprasop, K., & Kerdprasop, N. (2021). The Study of Noise Effect On CNN-Based Deep Learning from Medical Images. *Int.J.Mach.Learn.Comput*, 11, 202–207.
- Streiner, D. L., & Cairney, J. (2007). What's Under The ROC? An Introduction to Receiver Operating Characteristics Curves. *The Canadian Journal of Psychiatry*, 52(2), 121–128.
- Suliman, A., & Zhang, Y. (2015). A Review on Back-Propagation Neural Networks in the Application of Remote Sensing Image Classification. *Journal Of Earth Science and Engineering*, 5(1), 52–65.
- Suparta, W., & Samah, A. A. (2020). *Rainfall Prediction by Using ANFIS Times Series Technique in South Tangerang, Indonesia*. Elsevier BV. 10.1016/J.Geog.2020.08.001
- Tai, X., Liu, H., Chan, R. H., & Li, L. (2024). A Mathematical Explanation of Unet. *Arxiv Preprint Arxiv:2410.04434*,
- Takahashi, S., Sakaguchi, Y., Kouno, N., Takasawa, K., Ishizu, K., Akagi, Y., Aoyama, R., Teraya, N., Bolatkan, A., Shinkai, N., Machino, H., Kobayashi, K., Asada, K., Komatsu, M., Kaneko, S., Sugiyama, M., & Hamamoto, R. (2024). Comparison Of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review. *Journal Of Medical Systems*, 48(1)10.1007/S10916-024-02105-8
- Talpur, N., Abdulkadir, S. J., Alhussian, H., Hasan, M. H., Aziz, N., & Bamhdi, A. (2022). *Deep Neuro-Fuzzy System Application Trends, Challenges, And Future Perspectives: A Systematic Survey*. Springer Science and Business Media LLC. 10.1007/S10462-022-10188-3

- Tanimola, O., Shobayo, O., Popoola, O., & Okoyeigbo, O. (2024). Breast Cancer Classification Using Fine-Tuned SWIN Transformer Model on Mammographic Images. *Analytics*, 3(4), 461. 10.3390/Analytics3040026
- Tanzi, L., Audisio, A., Cirrincione, G., Aprato, A., & Vezzetti, E. (2022). Vision Transformer for Femur Fracture Classification. *Injury*, 53(7), 2625–2634.
- Tian, D., Jiang, S., Zhang, L., Lu, X., & Xu, Y. (2023). The Role of Large Language Models in Medical Image Processing: A Narrative Review. *Quantitative Imaging in Medicine and Surgery*, 14(1), 1108. 10.21037/Qims-23-892
- Tian, Y., & Zhang, Y. (2022). A Comprehensive Survey on Regularization Strategies in Machine Learning. *Information Fusion*, 80, 146–166.
- Tiwari, R. G., Misra, A., Maheshwari, S., Gautam, V., Sharma, P., & Trivedi, N. K. (2025). Adaptive Neuro-FUZZY Inference System-Fusion-Deep Belief Network for Brain Tumor Detection Using MRI Images with Feature Extraction. *Biomedical Signal Processing and Control*, 103, 107387.
- Tomlinson, R. E., & Silva, M. J. (2013). Skeletal Blood Flow in Bone Repair and Maintenance. *Bone Research*, 1(1), 311–322.
- Touati, R., Le, W. T., & Kadoury, S. (2021). A Feature Invariant Generative Adversarial Network for Head and Neck MRI/CT Image Synthesis. *IOP Publishing*. 10.1088/1361-6560/Abf1bb
- Tsiridis, E., Upadhyay, N., & Giannoudis, P. (2007). Molecular Aspects of Fracture Healing: Which Are the Important Molecules? *Injury*, 38(1), S11–S25.
- Übeyli, E. D. (2009). Adaptive Neuro-Fuzzy Inference Systems for Automatic Detection of Breast Cancer. *Journal Of Medical Systems*, 33(5), 353.
- Usman, M., Evans, R., Saatchi, R., Kingshott, R., & Elphick, H. (2019). Non-invasive respiration monitoring by thermal imaging to detect sleep apnoea. In: The 32nd International Congress and Exhibition on Condition Monitoring and Diagnostic Engineering Management, Huddersfield, 3 Sep 2019 - 5 Sep 2019.
- Usamentiaga, R., Venegas, P., Guerediaga, J., Vega, L., Molleda, J., & Bulnes, F. G. (2014). Infrared Thermography for Temperature Measurement and Non-Destructive Testing. *Sensors*, 14(7), 12305–12348.

- Uzunova, H., Ehrhardt, J., & Handels, H. (2020). Memory-Efficient GAN-Based Domain Translation of High-Resolution 3D Medical Images. *Computerized Medical Imaging and Graphics*, 8610.1016/J.Compmedimag.2020.101801
- Van, M., Verma, P., & Wu, X. (2024). On Large Visual Language Models for Medical Imaging Analysis: An Empirical Study. Paper Presented at the *2024 IEEE/ACM Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 172–176.
- Vardasca, Magalhaes, & Mendes. (2019). Biomedical Applications of Infrared Thermal Imaging: Current State of Machine Learning Classification. *Proceedings*, 27(1), 46. 10.3390/Proceedings2019027046
- Vasile, C. M., Udriștoiu, A. L., Ghenea, A. E., Popescu, M., Gheonea, C., Niculescu, C. E., Ungureanu, A. M., Udriștoiu, Ș, Drocaș, A. I., & Gruionu, L. G. (2021). Intelligent Diagnosis of Thyroid Ultrasound Imaging Using an Ensemble of Deep Learning Methods. *Medicina*, 57(4), 395.
- Venkatesan, R., & Li, B. (2017). *Convolutional Neural Networks in Visual Computing: A Concise Guide*. CRC Press.
- Vorontsov, E., Molchanov, P., Gazda, M., Beckham, C., Kautz, J., & Kadoury, S. (2022). Towards Annotation-Efficient Segmentation Via Image-To-Image Translation. *Medical Image Analysis*, 8210.1016/J.Media.2022.102624
- Vuksanović, V., Sheppard, L. W., & Stefanovska, A. (2008). Nonlinear Relationship Between Level of Blood Flow and Skin Temperature for Different Dynamics of Temperature Change. *Biophysical Journal*, 94(10), L78–L80.
- Walia, N., Singh, H., & Sharma, A. (2015). ANFIS: Adaptive Neuro-Fuzzy Inference System-A Survey. *International Journal of Computer Applications*, 123(13)
- Wallig, M. A., Bolon, B., Haschek, W. M., & Rousseaux, C. G. (2017). *Fundamentals Of Toxicologic Pathology*. Academic Press.
- Wang, C., Zhang, J., & Liu, S. (2022). Medical Ultrasound Image Segmentation with Deep Learning Models. *IEEE Access*, 11, 10158–10168.
- Wang, W., Huang, Y., Wang, Y., & Wang, L. (2014). Generalized Autoencoder: A Neural Network Framework for Dimensionality Reduction. Paper Presented at the *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 490–497.

- Wang, Y. (2020). A Mathematical Introduction to Generative Adversarial Nets (GAN). *Arxiv Preprint Arxiv:2009.00169*,
- Wang, Y., Jia, B., & Xian, C. (2023). Machine Learning and Unet Based Microfracture Evaluation from CT Images. *Geoenergy Science and Engineering*, 226, 211726.
- Wang, Z., Zou, Y., & Liu, P. X. (2021). Hybrid Dilation and Attention Residual U-Net for Medical Image Segmentation. *Computers In Biology and Medicine*, 13410.1016/J.Compbimed.2021.104449
- Wei, J., Wang, Q., Song, X., & Zhao, Z. (2023). The Status and Challenges of Image Data Augmentation Algorithms. Paper Presented at The *Journal of Physics: Conference Series*, 2456(1) 012041.
- Welling, R. D., Jacobson, J. A., Jamadar, D. A., Chong, S., Caoili, E. M., & Jebson, P. J. (2008). MDCT And Radiography of Wrist Fractures: Radiographic Sensitivity and Fracture Patterns. *American Journal of Roentgenology*, 190(1), 10–16.
- Willeminck, M. J., Koszek, W. A., Hardell, C., Wu, J., Fleischmann, D., Harvey, H., Folio, L. R., Summers, R. M., Rubin, D. L., & Lungren, M. P. (2020). Preparing Medical Imaging Data for Machine Learning. *Radiology*, 295(1), 4–15.
- Winsor, T., & Winsor, D. (1985). The Noninvasive Laboratory—History and Future of Thermography. *Angiology*, 36(6), 341–353.
- Wolf, D., Payer, T., Lisson, C. S., Lisson, C. G., Beer, M., Götz, M., & Ropinski, T. (2023). Self-Supervised Pre-Training with Contrastive and Masked Autoencoder Methods for Dealing with Small Datasets in Deep Learning for Medical Imaging. *Scientific Reports*, 13(1)10.1038/S41598-023-46433-0
- Wraighte, P. J., & Scammell, B. E. (2006). Principles Of Fracture Healing. *Surgery (Oxford)*, 24(6), 198–207.
- Wu, H., & Gu, X. (2015). Towards Dropout Training for Convolutional Neural Networks. *Neural Networks*, 71, 1–10.
- Wu, M., Wang, S., Pan, S., Terentis, A. C., Strasswimmer, J., & Zhu, X. (2021). Deep Learning Data Augmentation for Raman Spectroscopy Cancer Tissue Classification. *Scientific Reports*, 11(1), 23842.

- Wu, M., & Chen, L. (2015). Image Recognition Based on Deep Learning. Paper Presented at the 2015 Chinese Automation Congress (CAC), 542–546.
- Wu, Z., Shen, C., & Van Den Hengel, A. (2019). Wider or Deeper: Revisiting the Resnet Model for Visual Recognition. *Pattern Recognition*, 90, 119. 10.1016/J.Patcog.2019.01.006
- Xiao, M., Wu, Y., Zuo, G., Fan, S., Yu, H., Shaikh, Z. A., & Wen, Z. (2021). Addressing Overfitting Problem in Deep Learning-Based Solutions for Next Generation Data-Driven Networks. *Wireless Communications and Mobile Computing*, 2021(1), 8493795.
- Xiao, Y., Chen, C., Wang, L., Yu, J., Fu, X., Zou, Y., Lin, Z., & Wang, K. (2023). A Novel Hybrid Generative Adversarial Network for CT And MRI Super-Resolution Reconstruction. *Physics In Medicine & Biology*, 68(13)10.1088/1361-6560/Acdc7e
- Xu, B., & Yang, G. (2025). Interpretability Research of Deep Learning: A Literature Survey. *Information Fusion*, 115, 102721.
- Xu, M., Yoon, S., Fuentes, A., & Park, D. S. (2023). A Comprehensive Survey of Image Augmentation Techniques for Deep Learning. *Pattern Recognition*, 137, 109347.
- Xun, S., Li, D., Zhu, H., Chen, M., Wang, J., Li, J., Chen, M., Wu, B., Zhang, H., Chai, X., Jiang, Z., Zhang, Y., & Huang, P. (2021). Generative Adversarial Networks in Medical Image Segmentation: A Review. *Computers In Biology and Medicine*, 14010.1016/J.Compbimed.2021.105063
- Yadav, D. P., Sharma, A., Athithan, S., Bholra, A., Sharma, B., & Dhaou, I. B. (2022). Hybrid Sfnnet Model for Bone Fracture Detection and Classification Using ML/DL. *Sensors*, 22(15), 5823.
- Yan, L., Ling, S., Mao, R., Xi, H., & Wang, F. (2024). A Deep Learning Framework for Identifying and Segmenting Three Vessels in Foetal Heart Ultrasound Images. *Biomedical Engineering Online*, 23(1), 39.
- Yang, D., Martinez, C., Visuña, L., Khandhar, H., Bhatt, C., & Carretero, J. (2021). Detection And Analysis Of COVID-19 In Medical Images Using Deep Learning Techniques. *Scientific Reports*, 11(1)10.1038/S41598-021-99015-3
- Yang, Z., Sinnott, R. O., Bailey, J., & Ke, Q. (2023). A Survey of Automated Data Augmentation Algorithms for Deep Learning-Based Image Classification Tasks. *Knowledge And Information Systems*, 65(7), 2805–2861.

- Yao, W., Bai, J., Liao, W., Chen, Y., Liu, M., & Xie, Y. (2024). From CNN To Transformer: A Review of Medical Image Segmentation Models. *Journal Of Imaging Informatics in Medicine*, 37(4), 1529. 10.1007/S10278-024-00981-7
- Yeom, C., & Kwak, K. (2018). *Performance Comparison of ANFIS Models by Input Space Partitioning Methods*. MDPI AG. 10.3390/Sym10120700
- Yi, J., Shin, Y., Hahn, S., & Lee, Y. H. (2022). Deep Learning Based Sarcopenia Prediction from Shear-Wave Ultrasonographic Elastography and Gray Scale Ultrasonography of Rectus Femoris Muscle. *Scientific Reports*, 12(1), 3596.
- Yi, X., Walia, E., & Babyn, P. (2019). Generative Adversarial Network in Medical Imaging: A Review. *Medical Image Analysis*, 5810.1016/J.Media.2019.101552
- Ying, X. (2019). An Overview of Overfitting and Its Solutions. Paper Presented at the *Journal of Physics: Conference Series*, 1168 022022.
- Yousefifard, M., Baikpour, M., Ghelichkhani, P., Asady, H., Darafarin, A., Esfahani, M. R. A., Hosseini, M., Yaseri, M., & Safari, S. (2016). Comparison Of Ultrasonography and Radiography in Detection of Thoracic Bone Fractures; A Systematic Review and Meta-Analysis. *Emergency*, 4(2), 55.
- Yu, X., Wang, J., Hong, Q., Teku, R., Wang, S., & Zhang, Y. (2022). Transfer Learning for Medical Images Analyses: A Survey. *Neurocomputing*, 489, 230. 10.1016/J.Neucom.2021.08.159
- Yu-Jen Chen, Y., Hua, K., Hsu, C., Cheng, W., & Hidayati, S. C. (2015). Computer-Aided Classification of Lung Nodules on Computed Tomography Images Via Deep Learning Technique. *Oncotargets And Therapy*, 10.2147/Ott.S80733
- Zhang, C., Bao, N., Sun, H., Li, H., Li, J., Qian, W., & Zhou, S. (2022). A Deep Learning Image Data Augmentation Method for Single Tumor Segmentation. *Frontiers In Oncology*, 12, 782988.
- Zhang, D., Islam, M. M., & Lu, G. (2012). A Review on Automatic Image Annotation Techniques. *Pattern Recognition*, 45(1), 346–362.
- Zhang, H., Guo, W., Zhang, S., Lu, H., & Zhao, X. (2022). Unsupervised Deep Anomaly Detection for Medical Images Using an Improved Adversarial Autoencoder. *Journal Of Digital Imaging*, 35(2), 153. 10.1007/S10278-021-00558-8

- Zhang, H., & Qie, Y. (2023). Applying Deep Learning to Medical Imaging: A Review. *Applied Sciences*, 13(18)10.3390/App131810521
- Zhang, H., Chen, J., Liao, B., Wu, F., & Bi, X. (2024). Deep Canonical Correlation Fusion Algorithm Based on Denoising Autoencoder for ASD Diagnosis and Pathogenic Brain Region Identification. *Interdisciplinary Sciences: Computational Life Sciences*, 16(2), 455. 10.1007/S12539-024-00625-Y
- Zhang, K., Sun, M., Han, T. X., Yuan, X., Guo, L., & Liu, T. (2017). Residual Networks of Residual Networks: Multilevel Residual Networks. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(6), 1303–1314.
- Zhang, R., Zhou, B., Lu, C., & Ma, M. (2022). The Performance Research of the Data Augmentation Method for Image Classification. *Mathematical Problems in Engineering*, 2022(1), 2964829.
- Zhang, Y. (2018). A Better Autoencoder for Image: Convolutional Autoencoder. Paper Presented at the *ICONIP17-DCEC*. Available Online: [Http://Users.Cecs.Anu.Edu.Au/Tom.Gedeon/Conf/Abcs2018/Paper/Abcs2018_Paper_58.Pdf](http://Users.Cecs.Anu.Edu.Au/Tom.Gedeon/Conf/Abcs2018/Paper/Abcs2018_Paper_58.Pdf) (Accessed On 23 March 2017), 34.
- Zhang, Y., Chen, J., Ma, X., Wang, G., Bhatti, U. A., & Huang, M. (2024). Interactive Medical Image Annotation Using Improved Attention U-Net with Compound Geodesic Distance. *Expert Systems with Applications*, 237, 121282.
- Zhao, K., Mei, Y., Wang, X., Ma, W., & Shen, W. (2025). Conditional GAN Performs Better Than Orthopedic Surgeon in Virtual Reduction of Femoral Neck Fracture. *BMC Musculoskeletal Disorders*, 26(1), 1–9.
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2020). Unet++: A Nested U-Net Architecture for Medical Image Segmentation. (Pp. 3). Springer International Publishing. 10.1007/978-3-030-00889-5_1
- Zhu, K., Chen, Y., Ouyang, X., White, G., & Agam, G. (2022). Fully RNN For Knee Ligament Tear Classification and Localization in MRI Scans. *Electronic Imaging*, 34(14) 10.2352/Ei.2022.34.14.Coimg-227
- Zhu, Q., Ye, H., Sun, L., Li, Z., Wang, R., Shi, F., Shen, D., & Zhang, D. (2021). GACDN: Generative Adversarial Feature Completion and Diagnosis Network For COVID-19. *BMC Medical Imaging*, 21(1)10.1186/S12880-021-00681-6

- Zhu, Q., He, Z., Zhang, T., & Cui, W. (2020). Improving Classification Performance of Softmax Loss Function Based on Scalable Batch-Normalization. *Applied Sciences*, 10(8), 2950.
- Zhu, Y., Meng, Z., Fan, X., Duan, Y., Jia, Y., Dong, T., Wang, Y., Song, J., Tian, J., & Wang, K. (2022). Deep Learning Radiomics of Dual-Modality Ultrasound Images for Hierarchical Diagnosis of Unexplained Cervical Lymphadenopathy. *BMC Medicine*, 20(1), 269.
- Zhu, Z. (2024). Advancements In Automated Classification of Chronic Obstructive Pulmonary Disease Based on Computed Tomography Imaging Features Through Deep Learning Approaches. *Respiratory Medicine*, 23410.1016/J.Rmed.2024.107809
- Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2020). A Comprehensive Survey on Transfer Learning. *Proceedings Of The IEEE*, 109(1), 43–76.

Appendices

i. Patient information for carer



Participant Information Sheet for Carer

Study title:

High Resolution Infrared Thermography as a Diagnostic Aid in Paediatric Wrist Injuries

1. Invitation to take part in research

Your child is being invited to take part in a research study that aims to develop thermal imaging to screen for wrist fracture. This information sheet explains the purpose of the study, how you and your child will be involved, your rights and how the data from the study is stored, processed and disseminated.

Please feel free to ask if there is anything that is not clear or if you would like more information.

2. What is the purpose of the study?

When assessing an injured child, doctors must decide whether or not there is an underlying bone fracture. Currently, the best way for doing this is to take an X-ray. However a significant proportion of X-rays show no fracture. The non-invasive, cost effective and harmless technique of thermal imaging holds promise as a technique that may reduce the number of X-rays taken to diagnose wrist fracture.

The aim of the study is to develop thermal imaging to screen for wrist fracture and exclude cases that are sprained and thus should not be X-rayed. We would like to test the accuracy of this new method on children attending the hospital for a wrist injury, with a suspicion of a fracture. For us to be able to assess the new method, we will only invite children that have an X-ray for their wrist injury.

3. Why has my child been chosen?

We need to compare the efficacy of the new (thermal imaging) against the current (X-ray) method. This requires recording from children who have wrist injury and the diagnosis (fracture or sprain) to be confirmed by X-ray. Your child is in this category and therefore he/she can help us with our study.

4. Does my child have to take part?

It is up to you and your child to decide whether or not to take part. You are both free to withdraw from the research at any time and without giving a reason. Your decisions about this will not affect the standard of care your child will receive. Irrespective of whether you and your child take part or not, the clinical treatment will be the same and the new method will not form part of your child's diagnosis.

Page 1 of 5

5. What will happen to my child if we agree take part?

If you are happy to take part, and are satisfied with the explanations from the research team, you will be invited to read and if you agree to sign a consent form. Your child will be given a simplified version of the Information Sheet and if he/she is happy to participate, will be invited to sign the assent form.

For the recording, you will be invited to go to a room where a thermal camera connected to a computer will be available to monitor your child's wrists. The thermal camera takes images from a distance about 1 meter to determine the amount of heat emitted from the skin above the wrists. Later on, the research team will process the images to investigate whether the approach can differentiate wrist fracture from sprain.

The camera will not be in direct contact with your child and will not cause any harm.

The recordings take about 5 minutes and the whole test takes about 15 minutes. After this you and your child will have no further involvement in the research study.

The recording will be stopped immediately, if you or your child wishes this.

After the recording, your child will be asked to either complete a simple evaluation questionnaire or these questions will be asked from him/her. These seek information about the comfort of the test.

You will be invited to stay with your child throughout the study.

6. How the test is carried out?

The child will be asked to:

- Put his/her wrists on a table and keep them there for about 30 seconds while some thermal images will be taken from them by a heat detecting camera.
- Place his/her wrists on a platform elevated at 45 degrees and a further 30 seconds thermal image recording is carried out.

7. What are the possible disadvantages and risks of taking part?

We very much appreciate your support of this study and we understand that you child can be in pain due to the injury. Participation in the test may increase your stay at the hospital for about 15-20 minutes. However, we will try our best to do the test while you are waiting for other things, to avoid any extra time spent in A&E. The risks associated with the x-ray are the same as undergoing normal clinical treatment.

8. What are the possible benefits of taking part?

There is no immediate benefit to your child from participation. But if the new method proved useful, it may reduce the number of X-ray scans taken in future thus reducing exposure to radiation as well as reducing NHS cost.

9. What will happen once the thermal imaging data are obtained?

The data will be carefully processed by the researchers to determine the accuracy of the method in screening for wrist fracture. If the method proves useful we aim to take it further toward further development.

10. Will my child's taking part in the research project be kept confidential?

Yes. We will follow ethical and legal practice and all information about your child will be handled in confidence. The data will be anonymised and the identity of the participants (you and your child) will not in any way revealed in any output (e.g. publications, presentations, web sites etc.)

11. What will happen if we don't want to carry on with the research?

If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.

12. What if there is a problem?

If you have any cause to complain about any aspect of the way in which you or your child has been approached or treated during the course of this study, the normal National Health Service complaints mechanisms are available to you and are not compromised in any way because you have taken part in a research study. If you have any complaints or concerns please contact

Dr Shammi Ramlakhan
Consultant Paediatric Emergency Physician
Sheffield Children's NHS Foundation Trust
Western Bank
Sheffield
S10 2TH

Telephones: 0114 271 7431, 0114 226 2331
Email: Shammi.Ramlakhan@sch.nhs.uk

13. Harm

In the unlikely event that taking part in this research project harms your child/ward, there is no special compensation arrangement. If your child/ward is harmed due to someone else's fault, then you may have ground for legal action, but you may have to pay your legal costs.

14. Will taking part in this study be kept confidential?

All information collected about your child during the course of the research will be kept strictly confidential. Identifiable data will be retained for 12 months after the end of the study. Any information about your child which leaves the hospital will be anonymised i.e. have their name and address removed so that your child cannot be identified from it. Once the study is complete all other information will be kept for up to 5 years and then destroyed in accordance with standard operating procedures.

Our procedures for handling, processing, storage and destruction of data are compliant with the Data Protection Act 2018.

The thermal images measurements will be stored on a secure password protected University computer. The data will be coded and only authorised people involved in the research study will have access.

Your child's medical notes may also be looked at by other medical staff within the hospital involved in the running and supervision of the study to check that it is being carried out correctly.

15. What will happen to the results of the research study?

We will interpret and analyse the recorded data with the aim of determining the efficacy of thermal imaging to screen for wrist fractures. If the method proves useful then we aim to further develop it toward a medical tool.

We also aim to publish the findings in various outputs such as research journals, books and present them in research meetings and conferences. The study could also appear on the Sheffield Children's Hospital or Sheffield Hallam University web sites.

The results will also appear in a dissertation of a Biomedical Sciences degree student. The student will assist in collecting and interpreting the data.

Your child's identity will not be recognised from the outputs.

16. Who is organising and funding the research?

The research is being organised by Sheffield Children's NHS Foundation Trust. The medical student will work on the study as part of her degree qualification.

17. Who has reviewed the study?

This study has been given a favourable ethical opinion for conduct in the NHS by the Sheffield Research Ethics Committee. It has also been approved by the Research Department at this hospital as well as by Sheffield Hallam University Ethics Committee.

Thank you for taking the time to read this information sheet

Page 4 of 5

The Sheffield Children's Hospital is the sponsor for this study based in the United Kingdom. We will be using information from your child in order to undertake this study and will act as the data controller for this study. This means that we are responsible for looking after your information and using it properly. The Sheffield Children's Hospital will keep identifiable information about your child for 1 year after the study has finished.

Your rights to access change or move your information are limited, as we need to manage your information in specific ways in order for the research to be reliable and accurate. If you withdraw from the study, we will keep the information about you that we have already obtained. To safeguard your rights, we will use the minimum personally-identifiable information possible.

You can find out more about how we use your information at <https://www.sheffieldchildrens.nhs.uk/research/research-fags/> or by contacting the Clinical Research Facility on 0114 271 7417.

The Sheffield Children's Hospital will collect information from you for this research study in accordance with our instructions.

The Sheffield Children's Hospital will use your name and contact details to contact you about the research study, and make sure that relevant information about the study is recorded for your care, and to oversee the quality of the study. Individuals from The Sheffield Children's Hospital and regulatory organisations may look at your medical and research records to check the accuracy of the research study. The Sheffield Children's Hospital will pass these details to The Sheffield Children's Hospital along with the information collected from you. The only people in The Sheffield Children's Hospital who will have access to information that identifies you will be people who need to contact you to provide the results of the study if you requested it or audit the data collection process. The people who analyse the information will not be able to identify you and will not be able to find out your name or contact details.

The Sheffield Children's Hospital will keep identifiable information about you from this study for 1 year after the study finishes.

ii. Patient information sheet for 5-year-old

Information Sheet for Participants Aged 5 Years

To be shown to and read by carer if required

Study title: Can a new method that measures heat detect broken bones?

1. Why are we doing this study?

We want to find out if a camera that can take photos of body heat can help doctors see broken bones.



2. Why did we invite you to take part?

We need to see how good the new way is in showing broken bones. It needs testing on some children with injured wrists. The test is comfortable.

3. What happens if I take part

We will take some photos of your injured wrists. This is done when your hand are on a table. Then we do this again when your hands are raised. This takes few minutes. Then you will be asked some questions about how you found the test.

4. Do I have to take part?

No. No one will be cross with you if you do not.

**Please feel free to ask any questions.
Thank you for reading this.**

iii. Patient information sheet for 6 – 12 years old



Information Sheet for Participants Aged 6-12 Years

To be shown to and read by parent/carer if required

Study title: Study to determine whether a heat detecting camera can help with identifying bone fractures.

1. Why are we doing this study?

We want to find out if a heat detecting camera can help doctors in identifying bone fractures.

2. Why me?

We need to test our new method that uses the heat detecting camera to see how well it can identify a bone fracture. As you have a wrist injury that may or may not be a bone fracture, you can help us with our study by taking part.

You do not have to accept this invitation if you do not want to.

3. Do I have to take part?

It is up to you to decide if you want to take part. No one will be cross with you if you say no.

You will receive the same medical treatment for your injury whether you take part or do not take part. The risks associated with the x-ray are the same as undergoing normal treatment.

4. What will happen to me if I take part?

You will go to a room where the heat detecting camera called a thermal camera is set up. The camera will not be in direct contact with you or cause harm. The recording takes about 5 minutes. The images will be stored and processed later.

After the recording, you will be invited to either complete a simple evaluation form to say how comfortable the test was. If you prefer we can ask you the questions and complete the form for you.

Your parent or carer will be able to stay with you throughout the study.

5. What will I be asked to do?

Thermal imaging developments for wrist fracture screening in children
Participant Information Sheet 6-12 years, Version 5.2, Date 11/02/2019 IRAS 253940

Page 1 of 2

You will be asked to:

- Put your wrists on a table and keep them there for about 30 seconds while some thermal images will be taken from them by a heat detecting camera.
- Place your wrists on a platform elevated at 45 degrees and a further 30 seconds thermal image recording is carried out.

You will not be asked to do anything else for the purposes of the research study.

6. Will taking part in the study benefit me?

Not on this occasion, but your help will enable us test our method. If the method is effective it could assist doctors to detect fractures in the future.

7. What happens when the research study stops?

The recordings will be processed by our research team to determine whether the new method can help doctors in identifying bone fractures.

8. What if I don't want to do the research anymore?

Please tell your carer or doctor if you do not wish to continue. They will not be cross with you. We will stop the recording. You will still have the same care whilst you are at hospital.

9. What if I wish to complain about the study?

If you want to complain you or carer can talk to Dr Shammi Ramlakhan at this hospital.

Dr Shammi Ramlakhan
Consultant Paediatric Emergency Physician
Sheffield Children's NHS Foundation Trust
Telephones: 0114 271 7431, 0114 226 2331
Email: Shammi.Ramlakhan@sch.nhs.uk

Thank you for reading this information sheet.

iv. **Patient information sheet for 13 – 15 years old.**



Information Sheet for Participants Aged 13-15 Years

Study title

High Resolution Infrared Thermography as a Diagnostic Aid in Paediatric Wrist Injuries

1. Invitation to take part in research

We would like to invite you to help us in our study. Please read this information sheet and talk to your parent or carer about the study. Ask us if there is anything that is not clear or if you want to know more. Take time to decide if you want to take part. It is up to you if you want to do this. If you don't then that's fine, you'll be looked after at the hospital just the same.

2. Why are we doing this study?

We wish to find out if the use of a camera that can measure the body's heat (called a thermal camera) can assist with detecting wrist fractures. If the method proves useful, it may reduce the number of X-rays taken in the future.

3. Why have I been asked to take part?

We need to investigate how well a thermal camera can assist in detecting wrist fractures. This requires evaluating the method on children who are having X-rays for their wrist injury.

4. Do I have to take part?

No, you decide whether to take part or not. It is completely up to you. You will be treated the same for your wrist injury irrespective of whether you do take part or not take part.

5. What will happen to me if I take part?

Your wrist has been examined by a doctor who has already decided you need an X-ray to determine whether you have bone fracture. Therefore, this study is not part of your examination. The risks associated with the x-ray are the same as undergoing normal clinical treatment.

If you take part in the research study, you will be taken to a room where a thermal camera is set up to measure the temperature of your wrists. The camera will not be in direct contact with you or cause harm. The recording takes about 5 minutes. The images will be stored and processed later.

After the recording, you will be invited to either complete a simple evaluation questionnaire or these questions will be asked of you. These seek information about the comfort of the test.

Your carer will be able to stay with you throughout the study.

6. What will I be asked to do?

You will be asked to:

- Put your wrists on a table and keep them there for about 30 seconds while some thermal images will be taken from them by a heat detecting camera.
- Place your wrists on a platform elevated at 45 degrees and a further 30 seconds thermal image recording is carried out.

You will not be asked to do anything else for the purposes of the research study.

7. Will the study benefit me?

The study will not benefit you but the information we get might help people with wrist injuries in the future.

8. What happens when the research study stops?

We will collect all the wrist images and process them to determine whether they can assist in detecting bone fractures.

9. What if I don't want to do the research anymore?

Just tell your carer, doctor or nurse at any time. They will not be cross with you. You will still have the same care whilst you are at hospital.

10. Can I be identified from the recorded data?

We will ensure your identity is not be revealed by the recorded data. When we publish our study in for example journals and books or present them in research conferences and web sites, we will not indicate your identity (name, address etc.).

Only the medical staff at the hospital who have access to your hospital record can relate your recorded data to your identity. For the purpose of this study, we will only record your wrists (not face).

11. What will happen to the results of the research study?

We would like others scientists and medical people know about the study's findings. So we aim to publish our finding (journal, books etc) and present them in research conferences. A medical student is also going to write a report (thesis) about this study.

12. Who is organising the research?

The research is being organised by Sheffield Children's NHS Foundation Trust.

13. Who has checked the study?

Before any research goes ahead it has to be checked by a Research Ethics Committee. This is a group of people who make sure that the research is safe to be carried out. This project has been approved by the Sheffield Ethics Committees.

14. What if there is a problem or something goes wrong?

We do not expect any problems or that anything will go wrong, but if does, tell us and we will try and sort it out straight away. You and your carer can either contact the project co-ordinator who is named in the next section.

15. Contact person

If you would like any further information about this study you could contact:

Dr Shammi Ramlakhan
Consultant Paediatric Emergency Physician
Sheffield Children's NHS Foundation Trust
Western Bank
Sheffield
S10 2TH

Telephones: 0114 271 7431, 0114 226 2331

Email: Shammi.Ramlakhan@sch.nhs.uk

Thank you for taking the time to read this

Please ask any questions if you need to.

v. Assent form



**ASSENT FORM FOR CHILDREN & YOUNG PEOPLE
(to be completed by the child/young person and their carer)**

Title of project:

High Resolution Infrared Thermography as a Diagnostic Aid in Paediatric Wrist Injuries

Participant study number:

Child (or if unable, carer on their behalf)/young person to please circle all they agree with:

- Have you read (or had read to you) about this project? Yes / No
- Has somebody else explained this project to you? Yes / No
- Do you understand what this project is about? Yes / No
- Have you asked all the questions you want? Yes / No
- Have you had your questions answered in a way you understand? Yes / No
- Do you understand that it's OK to stop taking part at any time? Yes / No
- Are you happy to take part? Yes / No

If any answers are 'No' or you don't want to take part, don't sign your name!

If you do want to take part, please write your name and today's date

Your name _____ Date _____

Your parent or carer must write their name here too if they are happy for you to take part

Name of Parent/Carer Date Signature

The researcher who explained this project to you needs to sign too:

Name of Researcher Date Signature

Thank you for your help. 1 copy for participant; 1 copy for researcher; 1 copy to be kept with hospital notes

vi. **Consent form**



Patient study number:

Carer Consent Form

Title of project:

High Resolution Infrared Thermography as a Diagnostic Aid in Paediatric Wrist Injuries

Names of researchers: Dr Shammi Ramlakhan, Ms Charlotte Reed, Professor Reza Saatchi

No.	Points	Please initial
1	I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions.	<input type="text"/>
2	I understand that my child's participation is voluntary and that I am free to withdraw my child at any time, without giving any reason, without my child's medical care or legal rights being affected.	<input type="text"/>
3	I understand that the relevant sections of any of my child's clinical record may be looked at by authorised researchers who are involved in the running of the study from Sheffield Children's NHS Foundation Trust where it is relevant to my child taking part in research. I give permission for Sheffield Children's NHS Foundation Trust staff to have access to my child's records.	<input type="text"/>
4	I agree thermal images of child's wrist to appear in publication and presentations.	<input type="text"/>
5	I agree to my child taking part in the above study.	<input type="text"/>

Name of Parent/Guardian

Date

Signature

Name of research team member
taking consent

Date

Signature

Authorised Researcher (Chief Investigator)

Date Signature

1 copy for parent; 1 copy for researcher; 1 copy to be kept with hospital notes

vii. **Epigeum ethics training certificate**



ix. Sample of patients' anonymised data

Patient Study Number	Continuous Variables				Sex (Male/Female)	Drug History (Paracetamol, Ibuprofen or Both)	Time since drugs taken (Hours, Minutes)
	Age (Years, Months)	Weight (kg)	Body Temperature (°C)	Time Since Injury (Days, Hours, Minutes)			
001	8y 7m	24.85	35.8	2h 35m	Female	Paracetamol 375mg Oral	at Triage (1h 19m ago)
002	8y 10m	50.0	N/A	22h 20m	Female	Paracetamol 700mg Oral	at Triage (29m ago)
003	11y 11m	67.8	36.6	48h 00m	Female	Paracetamol 1g Oral	at Triage (30m ago)
004	11y 6m	54.2	35.6	21h 05m	Female	Paracetamol 1g Oral	at Triage (30m ago)
005	15y 1m	60.25	35.3	2h 00m	Female	Paracetamol 1g Oral Ibuprofen 400mg Oral	at Triage (25m ago)
006	9y 9m	41.9	36.2	2h 00m	Male	Paracetamol 500mg Oral and Ibuprofen 200mg Oral	P at Home (1h 28m ago) I at Triage (53m ago)
007	10y 3m	55.0	N/A	9d 0h 0m	Female	Paracetamol 750mg Oral	at Triage (20m ago)
008	10y 5m	33.4	35.6	0h 30m	Female	Ibuprofen 500mg Oral	at Triage (23m ago)
009	9y 11m	35.3	36.4	1h 30m	Female	Paracetamol 625mg Oral	at Triage (38m ago)
010	10y 11m	30.05	36.2	22h 0m	Female	None	
011	9y 2m	32.3	36.3	2h 15m	Female	Ibuprofen 240mg Oral	at Triage (31m ago)
012	10y 10m	63.3	35.9	5d 6h 0m	Male	None	
013	11y 2m	51.4	35.0	2h 15m	Male	Paracetamol 500mg Oral and Ibuprofen 200mg Oral	at Home (2hr 55m ago)
014	5y 6m	25.1	35.8	3d 23h 50m	Male	None	
015	13y 2m	48.6	35.1	22h 45m	Male	None	
016	11y 9m	55.0	35.9	25h 00m	Female	None	
017	15y 0m	47.6	35.6	23h 45m	Female	Paracetamol 750mg Oral	at Triage (2h 29m ago)
018	12y 1m	41.1	35.8	2d 18h 30m	Male	Ibuprofen 200mg Oral	at Home (1h 00m ago)
019	7y 4m	25.5	35.9	7h 30m	Female	Ibuprofen 200mg Oral	at Home (2h 15m ago)
020	11y 11m	41.6	36.5	21h 00m	Female	Paracetamol 750mg Oral	at Triage (1h 22m ago)
021	13y 4m	56.6	35.5	4d 0h 0m	Male	None	
022	15y 3m	72.5	N/A	2d 23h 40m	Female	Paracetamol 500mg Oral	at Home (4h 35m ago)
023	14y 9m	N/A (Not triaged)	35.6	2d 2h 0m	Male	None	
024	15y 9m	55	36.3	2h 30m	Male	Ibuprofen 400mg Oral	at Triage (40m ago)
025	11y 6m	58.25	36.4	2h 15m	Male	Paracetamol 1g Oral	at Triage (0h 36m ago)
026	10y 8m	35.55	36.5	2h 35m	Female	Paracetamol 500mg Oral and Ibuprofen 200mg Oral	P at Home (1h 30m ago) I at Triage (26m ago)
027	10y 8m	37.7	35.9	2h 25m	Female	Paracetamol 625mg Oral	at Triage (0h 40m ago)
028	11y 3m	47.9	36.1	1h 00m	Male	Paracetamol 750mg Oral and Ibuprofen 200mg Oral	at Triage (0h 30m ago)
029	15y 0m	87.7	36.1	1h 05m	Male	Paracetamol 1gm Oral and Ibuprofen 400mg Oral	at Triage (0h 58m ago)
030	8y 2m	30.9	35.9	3h 50m	Female	Paracetamol 375mg Oral	at Home (2h 00m ago)
031	11y 0m	36.7	36.0	2h 00m	Female	Paracetamol 500mg Oral	at Home (1h 00m ago)
032	13y 3m	58.1	36.2	1h 30m	Male	Ibuprofen 400mg Oral	at Home (1hr 00m ago) at Triage (30m ago)
033	8y 6m	31.45	37.1	1d 22h 00m	Male	None	
034	14y 6m	49.4	35.9	21h 15m	Male	None	
035	5y 3m	18.3	36.4	1d 04h 05m	Female	Paracetamol 240mg Oral	at Home (2h 05m ago)
036	9y 6m	35.15	35.9	1d 09h 36m	Male	Paracetamol 625mg Oral	at Triage (1h 04m ago)
037	8y 9m	38.8	36.6	3h 5m	Female	None	
038	9y 11m	53.8	36.5	21h 45m	Female	Paracetamol 750mg Oral	at Triage (44m ago)
039	7y 3m	27.45	36.7	2h 45m	Male	None	
040	10y 5m	38.2	36.0	25h 10m	Male	Ibuprofen 300mg Oral	at Home (5h 40m ago)
041	11y 0m	49	35.9	2h 15m	Male	Paracetamol 500mg Oral	at Home (1h 45m ago)
042	10y 9m	49.5	36.1	22h 30m	Male	Paracetamol 500mg Oral	at Home (3h ago)
043	13y 1m	58.55	35.9	3h 30m	Male	Ibuprofen 400mg Oral and Paracetamol 1g Oral	P at Triage (1h 35m ago) I at Home (1h 45m ago)
044	12y 3m	66.8	36.5	5h 25m	Male	Paracetamol 1g Oral	at Triage (49m ago)
045	6y 11m	30.2	36.7	2d 3h 25m	Female	Paracetamol 500mg Oral	at triage (32m ago)
046	11y 4m	57.65	36.3	2h 45m	Male	Paracetamol 1g Oral	at Triage (37m ago)
047	10y 7m	54.6	36.9	2h 35m	Female	None	
048	11y 1m	34.85	35.9	2h 25m	Male	Paracetamol 625mg Oral and Ibuprofen 200mg Oral	at Triage (47m ago)
049	6y 3m	23.5	35.6	23h 05m	Male	Paracetamol 375mg Oral	at Triage (2h 7m ago)
050	9y 1m	38	36.5	3h 00m	Male	Paracetamol 625mg Oral and Ibuprofen 200mg Oral	at Triage (1h 57m ago)
051	12y 3m	80.3	35.6	4h 10m	Male	Paracetamol 500mg Oral and Ibuprofen 400mg Oral	P at Home (3h 10m ago) and I Triage(1h 54m ago)
052	10y 4m	40.8	36.6	2d 1h 50m	Female	None	
053	12y 5m	53	36.1	2h 00m	Female	None	
054	8y 2m	24.65	36.1	19h 05m	Male	Ibuprofen 300mg Oral	at Triage (N/A)

x. Camera details

FLIR T630sc Series Infrared Camera Specifications	
System Overview	
Detector Type	Uncooled <u>Microbolometer</u>
Spectral Range	5.75 - 13.0 μm
Resolution	640 x 480
Detector Pitch 25 μm	17 μm
NETD	<40 <u>mK</u>
Electronics/Imaging	
Time Constant	<8 <u>ms</u>
Frame Rate	30Hz
Dynamic Range	14-bit
Digital Data Streaming	Real-Time Radiometric = USB to PC Real-Time Non-Radiometric = MPEG via USB to PC
Analog Video	DVI over HDMI
GSP	Location data stores with every image
Command & Control	USB, <u>WiFi</u>
Measurement	
Standard Temperature Range	-40°C to 650°C
Accuracy	+/- 2°C +/- 2% (limited range)
Optics	
Camera f/#	f/1.0, Integrated Lens 18mm (25°)
Available Lenses	88.9mm (7°), 41.3mm (15°), 24.6mm (25°), 13.1mm (45°), 6.5mm (80°)
Close-up Lenses/ Microscopes	Close-up (25 μm), (50 μm), (100 μm)
Focus	Continuous Automatic or Manual (Motorized and tactile)
Image Presentation	
On-Camera Display	Touch Screen/4.3 in LCD Display (1024 x 600) LCD Viewfinder (800 x 600)
Automatic Gain Control	Manual, Linear, Histogram, DDE
Image Analysis	Spot Meters, Areas, Auto Hot/Cold Detection, Difference Temp, Isotherms, Alarms
Image Annotations	60 Sec Voice, Text, 4 x Markers, Sketch
Visible Image	5.0 Megapixel from Integrated Visible Camera
General	
Operating Temperature Range	-15°C to 50°C
Storage Temperature Range	-40°C to 70°C
Encapsulation	IP 54 (IEC 60529)
Bump/Vibration	25g (IEC 60068-2-29)/ 2g (IEC 60068-2-6)
Battery System	Li Ion, 4 Hours Operating Time
Weight w/Battery	1.3kg
Size (LxWxH)	143x195x95mm
Mounting	1/4"-20