VISION TRANSFORMER AND LANGUAGE MODEL-BASED RADIOLOGY REPORT GENERATION

Arfa Siddiqua
CSE Department
Shadan Women's College Of Engineering and Technology,
Hyderabad, India
arfasiddiqua29@gmail.com

Dr. K. Palani
CSE Department.
Shadan Women's College of Engineering and Technology,
Hyderabad, India
principalswcet2020@gmail.com

ABSTRACT

Recent advancements in machine learning have enabled the development of cutting-edge models for a variety of computer vision applications, including sequence prediction. A significant innovation in this field is the use of transformer-based models, originally popularized in natural language processing (NLP), for computer vision tasks. These transformer models have demonstrated strong performance in various tasks such as sentiment analysis, language translation, and caption generation. In the context of medical imaging, autonomous report generation has become increasingly vital for improving healthcare delivery. Traditionally, recurrent neural networks (RNNs) have been used as decoders to generate captions or reports based on encoded images, while convolutional neural networks (CNNs) have served as encoders for extracting spatial features from images. This paper proposes a novel approach that leverages pre-trained language transformers as decoders and vanilla image transformer architectures as encoders. To assess the efficacy of this approach, an ablation study is conducted using the Indiana University Chest X-Rays dataset. The results from the comparative analysis show that the proposed method significantly outperforms existing techniques, highlighting the potential of transformer-based models to revolutionize automated report generation in medical imaging and potentially extend to other domains.

I. INTRODUCTION

The creation of cutting-edge models for a variety of computer vision applications, including sequence prediction, has been made possible by recent developments in machine learning. A significant innovation has been the use of transformer-based models, which were first made prominent in natural language processing applications, to computer vision issues. These transformer-based models have shown to be quite effective at a variety of tasks, including sentiment analysis, language translation, and caption creation. Autonomous report production is becoming more necessary in the field of medical imaging, which can significantly affect the provision of healthcare. Historically, recurrent neural networks (RNN) have been used as decoders to produce captions or reports based on the encoded picture, whereas convolutional neural networks (CNN) have been used as encoders to extract spatial information from graphics.

The suggested approach uses a variety of pre-trained language transformers as decoders and pre-trained vanilla image transformer topologies as encoders. In order to evaluate the efficacy of the suggested technique across several assessment parameters, ablation research is carried out using the Indiana University Chest X-Rays dataset. According to the comparison study, the suggested methodology outperforms current methods by a significant margin, underscoring the potential of transformer-based models to change automated report production in medical imaging and maybe other fields. Train the whole network, transfer learning and classification. In this work, the trained networks Mobile Net are used to study their performance cardiovascular disease detection.

OBJECTIVE

With an emphasis on medical imaging and automated report synthesis, the project's goal is to assess and illustrate the efficacy of employing transformer architecture as both an encoder and a decoder in text or report writing activities. In addition to examining the potential of transformer-based decoders in producing precise and contextually relevant captions or reports from encoded image features, the project looks at how well transformer-based models capture spatial information from medical images in comparison to conventional convolutional neural network (CNN) encoders. Using the Indiana University Chest X-Rays dataset, the project also aims to run an ablation research in order to assess the effectiveness of the suggested technique using a number of assessment measures, including accuracy, precision, recall, and F1 score. To evaluate the possibility of the suggested method for improving the caliber and effectiveness of automated report production in medical imaging applications, the research will compare it with current CNN-RNN designs. In the end, the research seeks to illustrate the usefulness and influence of transformer architecture in healthcare and related fields by offering insightful information about the advantages and disadvantages of using it to caption or report writing tasks for medical imaging.

II. PROBLEM STATEMENT

The growing demand for automated and efficient medical report generation in healthcare necessitates the development of advanced models capable of generating accurate and comprehensive reports from medical images. Traditional approaches have relied on recurrent neural networks (RNNs) as decoders and convolutional neural networks (CNNs) as encoders for generating captions or reports based on image data. However, these methods have limitations in terms of accuracy, scalability, and interpretability. With the recent success of transformer-based models in natural language processing tasks such as language translation, sentiment analysis, and caption generation, there is a compelling opportunity to explore their application in medical image report generation.

EXISTING SYSTEM

When someone suggested using CNN-RNN architecture to create captions for photos, the process of creating medical reports began. But these findings were vague and overly basic. Attention was introduced and models like as RNN and CNN employed attention as more work was done in the field. Results were much better with this model. The invention of Transformers began. It is just attention-focused and devoid of convolution and repetition. Using several encoder and decoder layers and Multi Attention Self-attention (MSA), it outperformed earlier language models. Transformers are utilized not just with text but also with images, where they have shown to be more effective than many current methods for certain tasks

Disadvantage of Existing System

- Consistency was not maintained.
- Intricacy is high.

PROPOSED SYSTEM

The process of creating a diagnostic report is essentially an image-to-sequence issue with pixels as inputs. Finds, impressions, and tags make up a comprehensive diagnostic report. Prior approaches employ a multi-tier structure. Each chest x-ray picture has predicted labels generated by a multiclass classification using the tags as labels. The right description is applied following a semantic analysis of the image. The reports contain descriptions that are many sentences lengthy, and the quality and correctness of the report depend heavily on their development. To address this issue, numerous LSTM network-based solutions have been put forth, including. However, CNN is unable to encode all features in latent space, and report descriptions are lengthy sentences, which affects the accuracy of reports produced by LSTM.

Advantages of Proposed System

- LSTMs are excellent at recalling crucial details.
- Accurate machine training is possible;
- It can comprehend sentence context or forecast future values in data.
- Effective transfer learning and adaptability to new datasets or applications are made possible by it.

III. RELATED WORKS

Several research have investigated the application of deep learning models to medical picture processing and report production. Traditionally, convolutional neural networks (CNNs) were used to extract spatial characteristics from medical pictures, whilst recurrent neural networks (RNNs) were used as decoders to provide textual descriptions or reports. These techniques have proven effective in a variety of applications, including creating radiology reports from chest X-rays and CT scans. Zhang et al. (2017), for example, introduced an approach that combines CNNs for feature extraction with long short-term memory (LSTM) networks to generate medical captions, illustrating the possibility of merging image processing and language models for medical report production. However, these traditional approaches struggle to handle long-range relationships in the image-to-text production process.

IV. METHODOLOGY OF PROJECT

This project's approach consists of many important steps, starting with dataset preparation and progressing through model creation, pre-training, fine-tuning, and assessment. First, the Indiana University Chest X-Rays dataset is pre-processed, with pictures scaled and normalized, as well as medical reports cleaned and tokenized for model compatibility. The model architecture includes an image encoder (ViT) that extracts features from X-ray pictures, as well as a text decoder that generates medical reports based on the retrieved characteristics. The ViT model is fine-tuned on the dataset to adapt to the unique characteristics of medical imaging, whilst the language transformer is fine-tuned on medical reports to provide contextually relevant descriptions. Both models are originally pre-trained on large generic datasets to boost their generalization capabilities.

MODULE DESCRIPTION:

Data Collection:

In feature-based methods, various features of the website we have to collect the ECG images dataset. These features are then used to train deep learning models to detect cardiovascular diseases

Dataset:
In the first module, we developed the system to get the input dataset for the training and testing purpose. Dataset is given in the model folder. The dataset consists of 1377 ECG images.

Importing the necessary libraries:
We will be using Python language for this. First, we will import the necessary libraries such as keras for building the main model, sklearn for splitting the training and test data, PIL for converting the images into array of numbers and other libraries such as pandas, NumPy, matplotlib and TensorFlow.

Retrieving the images:
We will retrieve the images and their labels. Then resize the images to (224,224) as all images should have same size for recognition. Then convert the images into NumPy array

Splitting the dataset:
Split the dataset into train and test. 80% train data and 20% test data.

Building the model:
The concept of convolutional neural networks is very successful in image recognition. The key part to understand, which distinguishes CNN from traditional neural networks, is the convolution operation. Having an image at the input, CNN scans it many times to look for certain features and give the performance score depends on that score only predict
Person can have diseases or not.

Saving the Trained Model:
Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5.

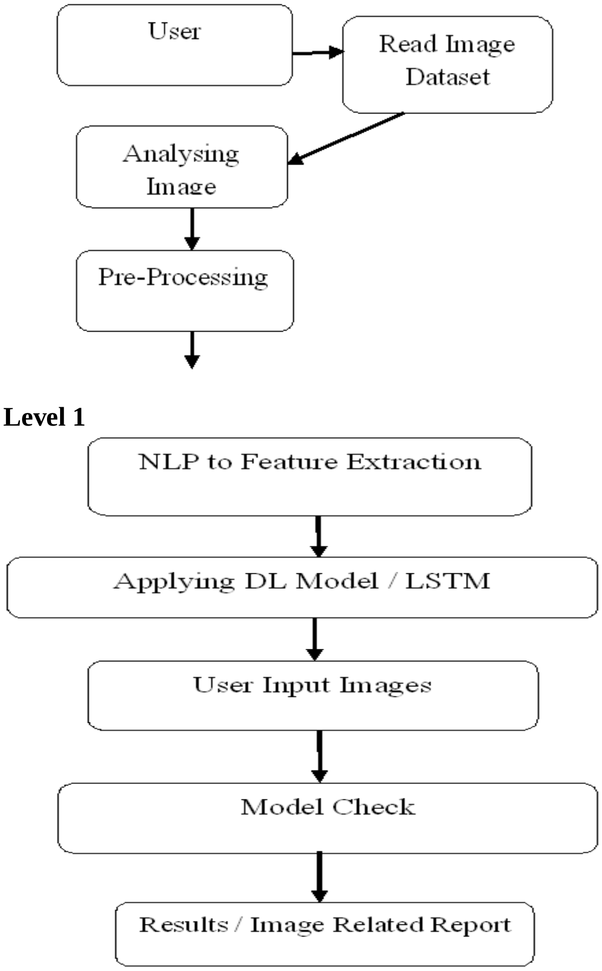
V. ALGORITHM USED IN PROJECT

Vision Transformer(LSTM)

The suggested system may learn hierarchical representations of chest X-ray images by including ViT, which captures both global context and fine-grained information that are essential for precise diagnosis and report production. In order to overcome CNNs' limitations in encoding full image features in the latent space, ViT's self-attention technique enables it to attend to pertinent picture regions and semantic information. Through this integration, the system may produce descriptions that are more contextually relevant and informative, increasing the overall accuracy and quality of diagnostic reports produced for medical imaging applications.

DATA FLOW DIAGRAM

DATA FLOW DIAGRAM
Level 0



Level 1

Fig5 Data Flow Diagram

VI. SYSTEM ARCHITECTURE

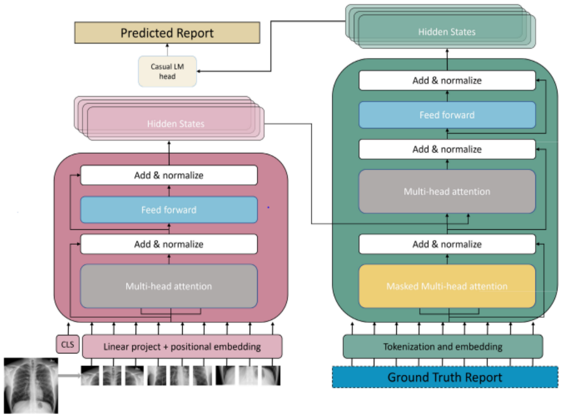
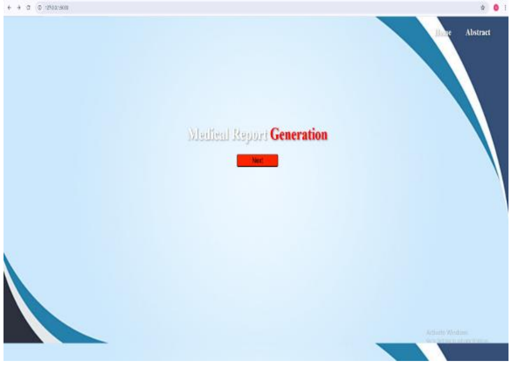
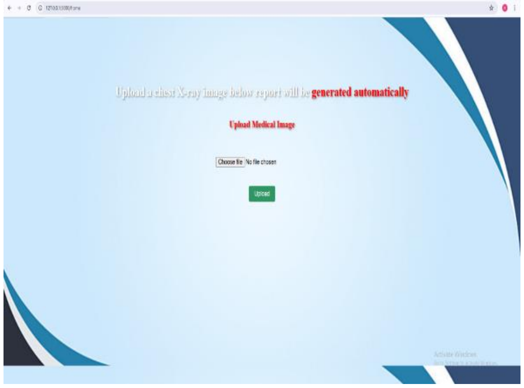


Fig6: System Architecture

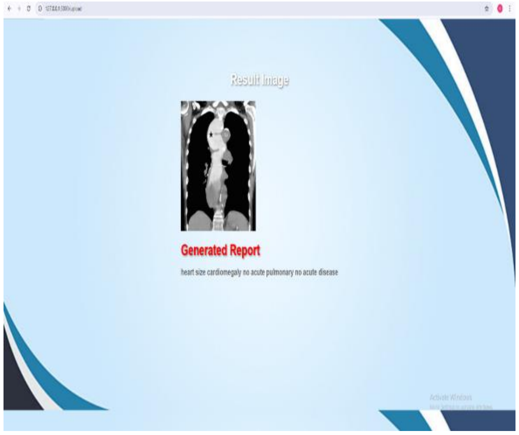
VII. RESULTS



7.1 Medical report generation


Upload a chest X-ray image below report will be generated automatically
Upload Medical Image
Choose file No file chosen
Upload

7.2 Upload image page



heart size cardiomegaly no acute pulmonary no acute disease

7.3 Generated medical reported image

VIII. FUTURE ENHANCEMENT

The system's capabilities and impact can be increased by focusing on a few crucial areas for future improvements in the chest X-ray image analysis and diagnostic report production system. First of all, a comprehensive patient profile can be created by combining multi-modal data sources such as medical history, patient demographics, and additional diagnostic testing. This improves diagnostic precision and allows for more individualized treatment. Clinical professionals can make better decisions by enhancing interpretability and emphasizing important aspects in images and generated reports by integrating attention processes into deep learning models like ViT and LSTM. Data annotation needs can be decreased by improving model performance and adaptation to domain-specific subtleties through the use of transfer learning and fine-tuning pre-trained models on particular medical datasets. It is possible to create interactive visualization tools that will make it easier for physicians and AI systems to collaborate and explore model outputs intuitively. Report content can be improved by developing NLP skills for semantic comprehension and contextual reasoning, and clinical workflows can benefit from real-time decision support integration, which provides instant insights. In conclusion, it is still critical to address ethical and regulatory issues, making sure that patient privacy, bias reduction, and model explainability are given top priority for the appropriate application of AI in healthcare.

IX. CONCLUSION

To sum up, the improvement that has been suggested for the system that analyzes chest X-ray images and generates diagnostic reports is intended to greatly expand its functionality and clinical application. Higher accuracy and greater adaptability to a range of patient profiles can be attained by the system through the integration of multi-modal data, the use of attention mechanisms, and the tuning of models through transfer learning. Enhancing the user experience and enabling smooth incorporation into clinical workflows are interactive visualization tools and real-time decision assistance elements. Deploying ethical and reliable AI solutions in healthcare also requires improving natural language processing (NLP) techniques for semantic comprehension and making sure that regulations and ethics are followed. All things considered, there is considerable potential for these upcoming developments to improve patient outcomes, diagnostic precision, and the general effectiveness of healthcare delivery.

REFERENCES

[1] "cardiovascular diseases," World Health Organization (WHO), 11. 06. 2021. [Online]. Available: https://www.who.int/health-topics/cardiovascular-diseases.

[2] "Common medical tests to diagnose heart conditions," Government of Westren Australia, Department of Health, [Online]. Available: https://www.healthywa.wa.gov.au/Articles/A_E/Common-medical-tests-to-diagnose-heart-conditions.

[3] M. Swathy and K. Saruladha, "A comparative study of classification and prediction of Cardio-vascular diseases (CVD) using Machine Learning and Deep Learning techniques," ICT Express, 2021. https://doi.org/10.1016/j.icte.2021.08.021.

[4] R. R. Lopes, H. Bleijendaal, L. A. Ramo, T. E. Verstraelen, A. S. Amin, A. A. Wilde, Y. M. Pinto, B. A. de Mol and H. A. Marquering, "Improving electrocardiogram-based detection of rare genetic heart disease using transfer learning: An application to phospholamban p. Arg14del mutation carriers," Computers in Biology and Medicine, vol. 131, no. 104262, 2021. https://doi.org/10.1016/j.compbiomed.2021.104262.

[5] R. J. Martis, U. R. Acharya and H. Adeli, "Current methods in electrocardiogram characterization," Computers in Biology and Medicine, vol. 48, pp. 133-149, 2014. https://doi.org/10.1016/j.compbiomed.2014.02.012.

[6] A. Rath, D. Mishra, G. Panda and S. C. Satapathy, "heart disease detection using deep learning methods from imbalanced ECG samples," Biomedical Signal Processing and Control, vol. 68, no. 102820, 2021. https://doi.org/10.1016/j.bspc.2021.102820.

[7] A. Mincholé and B. Rodriguez, "Artificial intelligence for the electrocardiogram," Nature Medicine, vol. 25, no. 1, pp. 22-23, 2019. https://doi.org/10.1038/s41591-018-0306-1.

[8] A. Isin and S. Ozdalili, "Cardiac arrhythmia detection using deep learning," Procedia Computer Science, vol. 120, pp. 268-275, 2017. https://doi.org/10.1016/j.procs.2017.11.238.

[9] H. Bleijendaal, L. A. Ramos, R. R. Lopes, T. E. Verstraelen, S. W. E. Baalman, M. D. Oudkerk Pool, F. V. Y. Tjong, F. M. Melgarejo-Meseguer, J. Gimeno-Blanes, J. R. Gimeno-Blanes, A. S. Amin, M. M. Winter, H. A. Marquering, W. E. M. Kok, A. H. Zwinderman, A. A. M. Wilde and Y. M. Pinto, "Computer versus Cardiologist: Is a machine learning algorithm able to outperform an expert in diagnosing phospholamban (PLN) p.Arg14del mutation on ECG?," Heart rhythm., vol. 18, no. 1, pp. 79-87, 2020. https://doi.org/10.1016/j.hrthm.2020.08.021.

[10] U. R. Acharya, H. Fujita, O. S. Lih, M. Adam, J. H. Tan and C. K. Chua, "Automated detection of coronary artery disease using different durations of ECG segments with convolutional neural network," Knowledge-Based Systems, vol. 132, pp. 62-71, 2017. https://doi.org/10.1016/j.knosys.2017.06.003.

[11] M. Kantardzic, Data Mining: Concepts, Models, Methods, and Algorithms, 3 ed., John Wiley & Sons, Inc., 2020.

[12] S. García, J. Luengo and F. Herrera, Data Preprocessing in Data Mining, 1 ed., Springer, 2015.

[13] G. Dougherty, Pattern Recognition and Classification: An Introduction, Springer, 2013.

[14] A. Subasi, Practical Machine Learning for Data Analysis Using Python, Academic Press, 2020.

[15] J. Soni, U. Ansari, D. Sharma and S. Soni, "Predictive data mining for medical diagnosis: An overview of heart disease prediction," International Journal of Computer Applications, vol. 17, no. 8, pp. 43-48, 2011.

[16] K. Dissanayake and M. G. Md Johar, "Comparative Study on Heart Disease Prediction Using Feature Selection Techniques on Classification Algorithms," Applied Computational Intelligence and Soft Computing, vol. 2021, 2021. https://doi.org/10.1155/2021/5581806.

[17] A. H. Gonsalves, F. Thabtah, R. M. A. Mohammad and G. Singh, "Prediction of coronary heart disease using machine learning: An experimental analysis," in Proceedings of the 2019 3rd International Conference on Deep Learning Technologies, 2019. https://doi.org/10.1145/3342999.3343015.

[18] H. Kim, M. I. M. Ishag, M. Piao, T. Kwon and K. H. Ryu, "A data mining approach for cardiovascular disease diagnosis using heart rate variability and images of carotid arteries," Symmetry, vol. 8, no. 6, 2016. https://doi.org/10.3390/sym8060047.

[19] T. Ozcan, "A new composite approach for COVID-19 detection in X-ray images," Applied Soft Computing, vol. 111, 2021. https://doi.org/10.1016/j.asoc.2021.107669.

[20] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size," arXiv preprint arXiv:1602.07360, 2016.

[21] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, vol. 25, pp. 1097-1105, 2012.

[22] A. H. Khan, M. Hussain and M. K. Malik, "Cardiac Disorder Classification by Electrocardiogram Sensing Using Deep Neural Network," Complexity, vol. 2021, 2021. https://doi.org/10.1155/2021/5512243.

[23] A. H. Khan and M. Hussain, "ECG Images dataset of Cardiac Patients," Mendeley Data, V2, 2021. https://doi.org/10.17632/gwbz3fsgp8.2.

[24] C. Potes, P. Saman, A. Rahman and B. Conroy, "Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds," in 2016 computing in cardiology conference (CinC), 2016.

[25] A. Nannavecchia, F. Girardi, P. R. Fina, M. Scalera and G. Dimauro, "Personal Heart Health Monitoring Based on 1D Convolutional Neural Network," Journal of Imaging, vol. 7, no. 2, 2021. https://doi.org/10.3390/jimaging7020026.

[26] Q. Zhang, D. Zhou and X. Zeng, "HeartID: A Multiresolution Convolutional Neural Network for ECG-Based Biometric Human Identification in Smart Health Applications," IEEE Access, vol. 5, pp. 11805-11816, 2017. https://doi.org/10.1109/ACCESS.2017.2707460.

[27] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," Computers in

biology and medicine, vol. 89, pp. 389-396, 2017. https://doi.org/10.1016/j.compbiomed.2017.08.022.

[28] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," Computational Intelligence and Neuroscience, vol. 2021, 2021. https://doi.org/10.1155/2021/8387680.

[29] P. Bizopoulos and D. Koutsouris, "Deep learning in cardiology," IEEE reviews in biomedical engineering, vol. 12, pp. 168-193, 2018. https://doi.org/10.1109/RBME.2018.2885714.

[30] S. Kiranyaz, T. Ince and M. Gabbouj, "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks," IEEE Transactions on Biomedical Engineering, vol. 63, no. 3, pp. 664-675, 2016. https://doi.org/10.1109/TBME.2015.2468589.

[31] M. Zubair, J. Kim and. C. Yoon, "An Automated ECG Beat Classification System Using Convolutional Neural Networks," in 2016 6th International Conference on IT Convergence and Security (ICITCS), 2016. https://doi.org/10.1109/ICITCS.2016.7740310.

[32] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018. arXiv:1801.04381.

[33] T. Rahman, A. Akinbi, M. E. Chowdhury, T. A. Rashid, A. Şengür, A. Khandakar, K. R. Islam and A. M. Ismael, "COV-ECGNET: COVID-19 detection using ECG trace images with deep convolutional neural network," 2021. arXiv preprint arXiv:2106.00436.

[34] G. Huang, Z. Liu, L. V. D. Maaten and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.

[35] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[36] A. Pal, R. Srivastva and Y. N. Singh, "CardioNet: An Efficient ECG Arrhythmia Classification System Using Transfer Learning," Big Data Research, vol. 26, p. 100271, 2021. https://doi.org/10.1016/j.bdr.2021.100271.

[37] R. Avanzato and F. Beritelli, "Automatic ECG diagnosis using convolutional neural network," Electronics, vol. 9, no. 6, p. 951, 2020. https://doi.org/10.3390/electronics9060951.

[38] U. R. Acharya, H. Fujita, S. L. Oh, Y. Hagiwara, J. H. Tan and M. Adam, "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals," Information Sciences, vol. 415–416, pp. 190-198, 2017. https://doi.org/10.1016/j.ins.2017.06.027.

[39] M. Naz, J. H. Shah, M. A. Khan, M. Sharif, M. Raza, and R. Damaševičius, “From ECG signals to images: a transformation-based approach for deep learning,” Peerj Comput Sci, vol. 7, p. e386, 2021, doi: 10.7717/peerj-cs.386.

[40] H. El-Amir and M. Hamdy, Deep Learning Pipeline: Building a Deep Learning Model with TensorFlow, Apress Media, 2020.

[41]. M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt and B. Scholkopf, "Support vector machines," IEEE Intelligent Systems and their Applications, vol. 13, no. 4, pp. 18-28, 1998. https://doi.org/10.1109/5254.708428.

[42] M. Abubaker and W. M. Ashour, "Efficient data clustering algorithms: improvements over Kmeans," International Journal of Intelligent Systems and Applications, vol. 5, no. 3, pp. 37-49, 2013. https://doi.org/10.5815/ijisa.2013.03.04.

[43] B. Charbuty and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," Journal of Applied Science and Technology Trends, vol. 2, no. 1, pp. 20-28, 2021.

[44] L. Bierman , "Random forests," Machine learning, vol. 45, no. 1, pp. 5-32, 2001. https://doi.org/10.1023/A:1010933404324.

[45] E. Miranda, E. Irwansyah,, A. Y. Amelga, M. M. Maribondang and M. Salim, "Detection of cardiovascular disease risk's level for adults using naive Bayes classifier," Healthcare informatics research, vol. 22, no. 3, pp. 196-205, 2016. https://doi.org/10.4258/hir.2016.22.3.196.

[46] G. Masetti and F. D. Giandomenico, "Analyzing Forward Robustness of Feedforward Deep Neural Networks with LeakyReLU Activation Function Through Symbolic Propagation," in Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2020.

[47] S. Shahinfar, P. Meek, and G. Falzon, "How many images do I need?" Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring, " Ecological Informatics, Vol. 57, 101085, 2020, https://doi.org/10.1016/j.ecoinf.2020.101085.

[48] B. Zoph, E. D. Cubuk, G. Ghiasi, T. Lin, J. Shlens, and Q. V. Le. "Learning data augmentation strategies for object detection." In European conference on computer vision, pp. 566-583. Springer, Cham, 2020.