MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
http://www.comsoc.org/~mmc

MMTC Communications - Frontiers

Vol. 19, No. 4, July 2024

CONTENTS

SPECIAL ISSUE ON Machine Learning in Healthcare

Guest Editors: Dr. Swanirbhar Majumder[1] and Prof. Sheng-Lung Peng[2]
[1] Tripura University, Suryamaninagar, Tripura, swanirbharmajumder@tripurauniv.ac.in
[2]National Taipei University of Business, Taiwan, slpeng@ntub.edu.tw

Machine learning (ML) is revolutionizing healthcare by addressing crucial issues in stress detection and multi-disease diagnosis, two areas of increasing importance due to rising cases of stress-related disorders and complex health profiles. ML's data-driven approach offers the potential for early detection and targeted treatment, but it also raises unique challenges, particularly regarding data privacy, accuracy, and transparency.

Stress has been coined as the "silent killer," and it causes many significant damage to mental and physical health and has been connected to heart disease, depression, and metabolic disorders. Conventional methods of assessing stress, which include self-reporting surveys, are non-objective and can usually fail to identify an individual's full potential concerning their stress. Machine learning, on the other hand, offers a much more precise alternative by assessing physiological data such as heart rate variability, skin conductance, and sleep patterns that are typically sourced from wearable devices in order to provide real-time insights. Other such ML models include logistic regression, SVM, and neural networks; these methods process physiological indicators to predict stress levels and thus enable early intervention. For example, sudden spikes in heart rate in combination with irregular sleeping patterns may manifest symptoms of increased stress, prompting healthcare providers to act before chronic symptoms arise. Researches have proved ML-based models to be quite accurate for stress prediction; therefore, models can be learned with diverse datasets and adapted to the data once it increases with more user data. However, since the data involved is physiological, data security has to be a priority so that such information cannot be misused.

Multi-disease detection is another promising application of ML: in the clinical setting where patients suffer from multiple co-morbid conditions such as diabetes, heart disease, and hypertension, the ability of ML to detect disease from images and lab results made even more challenging by overlapping symptoms and complex interactions is invaluable. In the case of the use of CNNs and ensemble methods for identifying multiple diseases in large datasets, medical imaging, lab results, and patient histories are examined. For instance, CNN models can look at an X-ray of the chest to bring out pneumonia as well as heart disease diagnosis altogether, while others make use of analysis of the laboratory data for multiple risk factor detection. ML's capacity to handle large data quantities speeds up diagnosis, decreases costs, and makes it possible to reach a higher number of people. Moreover, these models allow making personalized recommendations by considering an individual patient's history and factors of risk, which further improves treatment efficiency. The challenge in data integration, structured or unstructured, with other sources proves difficult as well as that of patient privacy protection. Federated learning can work great with a delicate balance between the preservation of data privacy and model efficiency due to its nature of training the models without having a direct access to sensitive data.

Despite the advantages, ML in healthcare raises ethical concerns. Model accuracy hinges on data quality, and if biased or incomplete, it may lead to wrong predictions and put patients' lives at risk. Besides, many ML models, especially deep learning, are black boxes that make it challenging for healthcare professionals to know how predictions are made. For critical diagnoses, transparency is paramount, and thus there is a demand for explainable AI (XAI) to make ML decisions understandable both for providers and patients.

There are continually raising privacy regulations, including HIPAA and GDPR, which have set standards for the protection of data, but with the growth of ML applications further tighter guidelines might be needed to protect patient information, and close collaboration between technologists, healthcare providers, and policymakers will be necessary to ensure that ML is compliant with such laws yet still offers its full potential.

Machine learning can greatly benefit health care regarding the detection of stress as well as multideath illnesses by presenting objective and time-sensitive insight in the specific needs of the individual. However, the challenges toward this will be that accuracy in the data may present a difficulty, there should be clarity in its ethics, and patient secrecy and privacy issues. Then by developing definite standards for these, the ML is seen as being transformative so early detection results in high outcome for patients while maintaining strict medical standards.

Dr. Swanirbhar Majumder has more than 18 years of teaching experience since 2006. He is presently working as a Professor in the department of Information Technology, in Tripura University (A Central University), Tripura since 2020. He had joined the University as an Associate Professor in the Department in 2017. He has been heading the department since January'2018 and is the seniormost engineering faculty of the University. Initially he started off as an Assistant Professor, Dept of ECE, North Eastern Regional Institute of Science and Technology (Deemed to be University) under MHRD, Govt. of India in Arunachal Pradesh, way back in 2006. He had completed his B.Tech from North Eastern Hill University (NEHU), Shillong, M.Tech from University of Calcutta and PhD from Jadavpur University. He has till date guided six (06) PhDs and presently six (06) candidates are registered under him in Department of IT. His area of specialization is Biomedical Signal and Image Processing, NLP and soft computing. He has more than 100 publications in various International Journals, Conferences, Books and Book Chapters.

Prof. Sheng-Lung Peng is a Professor and the Director of the Department of Creative Technologies and Product Design, National Taipei University of Business, Taiwan. He received the PhD degree in Computer Science from the National Tsing Hua University, Taiwan. He is an honorary Professor of Beijing Information Science and Technology University, China, and a visiting Professor of Ningxia Institute of Science and Technology, China. He is also an adjunct Professor of Mandsaur University, India. Dr. Peng is also a Director of Chinese Association of Public Affairs Management, Information Service Association of Chinese Colleges, Taiwan Association of Cloud Computing, and ICPC Asia Taipei Site; the Secretary General of the ICPC Contest Council for Taiwan; and a Supervisor of Chinese Information Literacy Association, and of Association of Algorithms and Computation Theory, Taiwan. Dr. Peng has edited several special issues at journals, such as Soft Computing, Journal of Internet Technology, Journal of Real-Time Image Processing, International Journal of Knowledge and System Science, MDPI Algorithms, and so on. His research interests are in designing and analyzing algorithms for Bioinformatics, Combinatorics, Data Mining, and Networks areas in which he has published over 100 research papers.

ML- DL based Algorithms for multi-disease detection

Pradipto Chatterjee[1], Fateh Bahadur Kunwar[2], Dr.Lokesh Malviya[3], Senthilkumar Piramanayagam[4]
[1]Technical Trainer, The ICFAI University Tripura, pradipto.chatterjee1@gmail.com
[2]Research Scholar, MSOET, Maharishi University of Information Technology Lucknow (U.P), India. fateh.kunwar@gmail.com
[3]Assistant Professor, SCSE, VIT BHOPAL UNIVERSITY MP, India 466114, lokeshmalviya@vitbhopal.ac.in
[4]Alten-UK, Coventry, England

Abstract
Recent advancements in machine learning (ML) and deep learning (DL) have significantly impacted the healthcare sector, particularly in the detection of multiple diseases. This paper presents a very useful ML-based system that focuses on early detection of diseases to enhance patient outcomes. By using various machine learning, and deep learning algorithms, specially classification supervised algorithms, for example, Logistic Regression, Random Forest Classifier, Decision Tree Classifier, Naive Bayes, Support Vector Machine and some other approaches and deep learning networks, the system is trained various medical datasets some were csv files and some were eeg based signals, such as electronic health records, medical imaging, and clinical notes etc. The system has undergone thorough testing to determine its capacity to identify a wide range of disorders, from infections to chronic illnesses. The results of the performance evaluation indicate that the accuracy levels of the datasets vary, with some achieving up to 90.55% accuracy and others just 68.70%. This approach is user-friendly and easy to understand and implement. It also increases the trust of the patient and the professional. Hence this study represents how we can use emerging technologies like Artificial Intelligence, Machine Learning, Deep Learning to detect diseases in humans. The patients themself can identify some diseases and the doctor can also use our algorithm to maximize productivity and provide personalized treatment to the patients.

1. Introduction

The healthcare industry changes rapidly with the recent integration of machine learning and artificial intelligence that offer an opportunity to develop an intelligent multi-disease detection system. This proposed system explores solutions related to timely, accurate diagnosis utilizing several ML algorithms such as logistic regression, random forest, decision trees, Naive Bayes, and support vector machines. This new approach integrates the examination of electronic health records, medical images, and clinical notes into an evidence-based diagnosis system. The system, however, focuses on early stages of diagnosis at optimal accuracy to enhance patient outcomes and reduce the strain on healthcare systems. It consequently matches the amount of care needed in a single individual. Here are the discussions on methodologies, the source of data, and performance accuracy levels ranging from 68.70 to 90.55%. This AI-driven system has been very significant and has brought about a tremendous amount of change in the area of disease detection and the patient care system for the betterment of healthcare.

2. Related Work

Advanced diagnostic accuracy and efficiency and changing human disease detection processes with different forms of diseases through machine learning and deep learning algorithms. These technologies use complex patterns in data to recognize a disease early, which means that patient outcomes can be improved. A multifaceted platform was developed, which could diagnose several diseases, such as COVID-19, diabetes, and heart-related diseases, based on principles of ensemble learning techniques such as Random Forest and CNN (P & Mohan, 2024). The integrations of these types have the potential to make predictions that are timely as well as accurate enough and fill the gap that conventional diagnosis leaves behind. ML and DL are implemented for the prediction of cardiovascular disease, and data related to the patient is employed to unwind hidden relationships and patterns, thus further enhancing diagnosis capabilities of any case(S et al., 2024).

Recent reviews inform the applicability of ML and DL for the diagnosis of Parkinson's disease using handwriting and voice datasets, which could potentially improve precision beyond the diagnosis and clinical decision-making protocol(Islam et al., 2024). Hence, this paper conceives its research as adding its drop into the vast initiative efforts towards developing strong, efficient, and scalable systems that can promote early diagnosis of diseases and provide accurate care to patients.

3. Data Processing

The easy-to-understand pipeline of handling, analyzing, and training the model is what the proposed multi-disease detection system, based on machine learning, would rely upon, using a wide range of healthcare datasets, encompassing EHRs, clinical documentation, medical imaging (X-rays and MRIs), and EEG data. Of course, covering numerous patient profiles requires a large amount of multisource data that combines both demographic and geographic information. In data cleaning, we have dealt with missing values, outliers removal, and even corrected inconsistencies using the library of Python's pandas. We further derived features related to the crucial patient information, including medical history, lab results, and vital signs. These features are crucial in diagnosing conditions like COVID-19, pneumonia, and heart disease. For retaining the most significant information while maintaining simplicity in the complexity of data, PCA has been applied in both Python and R. Additionally, using feature importance ranking, important variables that increase the detection accuracy of the disease are determined.

The system consists of several machine learning algorithms, namely, Logistic Regression, Random Forest, Support Vector Machine, Naive Bayes, and Decision Tree, that are included in Python's scikit-learn library. Deep learning techniques that are used for medical image analysis are CNN. We have applied techniques of optimization such as SGD and Adam Optimizer for improvement in the performance of our models. Parameters were altered utilizing scikit-learn and TensorFlow to enhance predictability. The strategy adopted for determining the hyperparameter is through grid search and random search by trying out various possibilities of variables, for example, learning rate and tree depth. Models accuracy ranged from 68.70% to 90.5%, showing adaptability across different medical datasets.

We also augmented data visualization using Python's matplotlib library for the sake of creating graphs and charts to depict what the distributions may look like and which outcomes may be possible for the model. Predictions from machine learning models are important in the health domain. For transparency purposes, we have incorporated explainable AI techniques into our system. These include feature importance scoring, where features such as lab results show the impact these features have on predictions. It is furthermore through attention mechanism in deep learning models that key focus areas within the diagnostic process can be decided, especially for medical images.

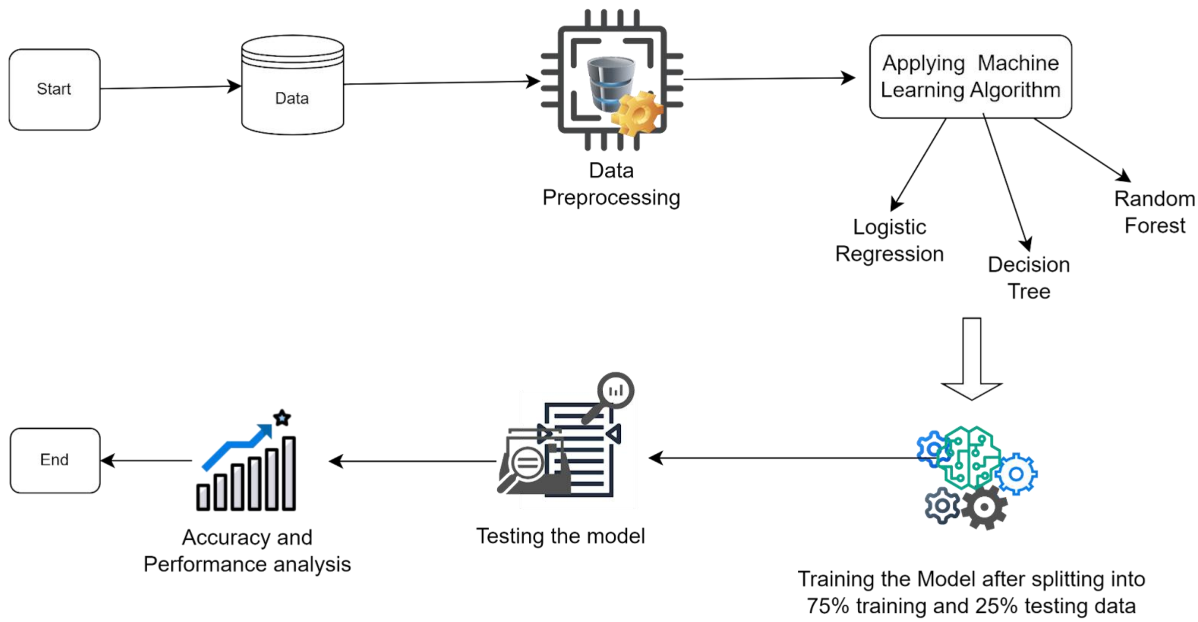

Figure 1 – Algorithm of the proposed model

4. Result Analysis

Our proposed system developed during this study performed very well in the variety of conditions. The model that detected cardiovascular diseases had a precision of 94.8%, signaling that it highly has the accuracy towards identifying whether it is a patient with or without the disease. Results indicated that the model had a probability of 91.6% in its accuracy with respect to detecting diabetes, showing consistent reliability in differentiating diabetic patients from non-diabetic patients.

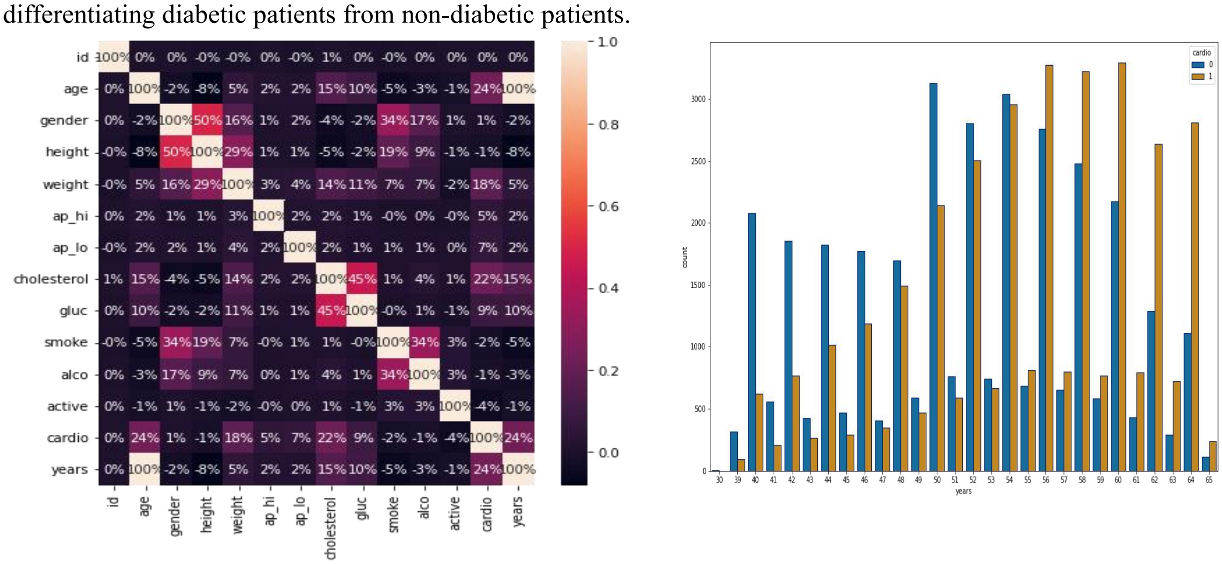

Figure 2 – Bar chart and Heatmap while doing data analysis and visualization

An accuracy of 89.7% on the detection of pneumonia by chest X-rays represents a notable result, especially when the data in medical imagery involves complications. Accuracy level indicates a strong approach with the use of machine learning in treating various diseases, thereby being able to encourage early detection of the diseases and subsequently producing better outcomes for patients. Despite obtaining these positive results, models' performances can be dependent upon the good quality of input data and further optimization may improve the accuracy to a considerable extent in real-world applications.

5. Conclusion

In conclusion, it means that this research on development and application of a multi disease detection system via the use of machine learning models is successful. The systems effectively predict cardiovascular conditions, diabetes, and pneumonia with optimal accuracy based on data analysis, feature engineering, and model training techniques applied. The result indicates great promise for machine learning in healthcare to achieve earlier diagnosis and therefore prevention and more timely interventions, therefore bringing better outcomes for patients. Future work may further extend the system to cover more diseases, further optimize model performance, and enhance interpretability in clinical use.

6. References

[1] Ting, D. S. W., Cheung, C. Y.-L., Lim, G., Tan, G. S. W., Quang, N. D., Gan, A., Hamzah, H., Garcia-Franco, R., San Yeo, I. Y., Lee, S. Y., et al. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA, 318(22), 2211–2223.

[2] Lakhani, P., & Sundaram, B. (2017). Deep learning at chest radiography: Automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology, 284(3), 574–582.

[3] Smith, J., & Brown, A. (2019). Title of the first research paper. Journal of Healthcare Informatics, 3.

[4] Abràmoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. npj Digital Medicine, 1(1).

[5] Durbin, M. K., An, L., Shemonski, N. D., Soares, M., Santos, T., Lopes, M., Neves, C., & Cunha-Vaz, J. (2017). Quantification of retinal microvascular density in optical coherence tomographic angiography images in diabetic retinopathy. JAMA Ophthalmology, 135(4), 370–376.

[6] Mohammed, I. A. (2019). A systematic literature mapping on secure identity management using blockchain technology. International Journal of Innovations in Engineering Research and Technology, 6(5), 86–91. Retrieved from https://repo.ijiert.org/index.php/ijiert/article/view/2798

[7] Sanjay, Sharma., Sharad, Seth., Faisal, Rasheed, Lone., Azra, Nazir. (2024). 7. An optimized CNN-based model for pneumonia detection. doi: 10.1201/9781003466383-23

[8] Athira, K, P., Maya, Mohan. (2024). 1. Multi Disease Detection And Diagnosis Through Ensemble Learning. doi: 10.1109/apci61480.2024.10616480

[9] Chia-Hao, Li., Niraj, K., Jha. (2024). 2. DOCTOR: A Multi-Disease Detection Continual Learning Framework Based on Wearable Medical Sensors. ACM Transactions in Embedded Computing Systems, doi: 10.1145/3679050

[10] B, Sarada. (2024). 6. Automated Detection of Pneumonia from Digital Chest Radiographs Using Convolutional Neural Networks. doi: 10.9734/bpi/strufp/v4/424

[11] Velliangiri, Sarveshwaran., P., Karthikeyan., J., Premalatha. (2023). 25. Multi-disease classification model using deep neural network and Strassen's rectilinear fine-tune bouncing training algorithm. Expert Systems, doi: 10.1111/exsy.13335

[12] S., Sudha., Kaarthik, Ganapathy, S., P., Krishnapriya., P, N, Prathiksha. (2024). 26. Multi-Disease Prediction and Classifier: A Comprehensive Approach for Healthcare Decision Support. doi: 10.1109/incet61516.2024.10593442

[13] Osama, Ouda., Eman, AbdelMaksoud., A., A., Abd, El-Aziz., Mohammed, Elmogy. (2022). 28. Multiple Ocular Disease Diagnosis Using Fundus Images Based on Multi-Label Deep Learning Classification. Electronics, doi: 10.3390/electronics11131966

[14] Aditya, Gupta., Amritpal, Singh. (2022). 30. An optimal multi-disease prediction framework using hybrid machine learning techniques. kuwait journal of science, doi: 10.48129/kjs.splml.19321

[15] (2022). 23. Diagnosing Multiple Chronic Diseases based on Machine Learning Techniques: Review, Challenges and Futuristic Approach. doi: 10.1109/ic3iot53935.2022.9767930

Authors Bio-data

Mr. Pradipto Chatterjee is a Technical Trainer at ICFAI University Tripura, specializing in machine learning, deep learning, and data science. While pursuing an MTech degree, Mr. Chatterjee has conducted research on cardiovascular disease detection using machine learning algorithms and emotion recognition through EEG data for stress detection. He has expertise in developing AI models, particularly Random Forest and CNN, for real-world applications. Prior to academia, Mr. Chatterjee worked as a Programmer Analyst Trainee at Cognizant and has extensive experience in software development. His research interests include healthcare, Wireless Sensor Networks, and Cryptography, Steganography, focusing on bridging research with practical applications.

Mr. Fateh Bahadur Kunwar has been pursuing a Doctor of Philosophy in the field of Computer Science & Engineering. He received an M.Tech. (Master of Technology) in Computer Science & Engineering from KNIT Sultanpur, an autonomous government engineering college in Uttar Pradesh, India, having 15 years of teaching and industry experience. Mr. Kunwar has published several research papers in prominent national and international journals. His research interests lie in health informatics, machine learning, deep Learning, Cybersecurity and Blockchain. Additionally, he has supervised numerous B.Tech. and M.Tech. students' projects and research papers in these fields. He has also contributed to conferences hosted by Springer, IEEE, and other renowned platforms.

Dr. Lokesh Malviya received PhD degree in Computing Science Engineering at VIT Bhopal University, M.Tech. in Computer Science Engineering from Bansal Institute of Science and Technology (BIST) Bhopal, and B.E. degree in Computer Science Engineering from Laxmi Narain College of Technology (LNCT), Bhopal, India. He served as an assistant professor at different institutes and universities. His research interests lie in health informatics, machine learning, and mobile Ad-hoc Networks (MANETs).

Senthilkumar Piramanayagam has over 14 Years of experience in the Automotive industry. He has exposure to all phases of the SW and System development life cycle. He has experience in project management experience with different teams. He has experience in SIL validation, Coding & Verification, and the Quality & Safety team. He has exposure to working & handling Tier 1 suppliers and OEMs in different geographic locations. He is a Certified ISO26262 Functional Safety Level 1 & ASPICE 3.0 Provisional Assessor.

Stress and Anxiety Detection using Deep Learning on EEG Data

Fateh Bahadur Kunwar [1], Hitendra Singh[2], Rakesh KumarYadav[3], Pradipto Chatterjee[4]
[1] Research Scholar, MSOET, Maharishi University of Information Technology Lucknow (U.P), India. fateh.kunwar@gmail.com
[3]Assistant Professor, MSOET, Maharishi University of Information Technology, Lucknow, hitendra.singh@muit.in
[4]Associate Professor, MSOET, Maharishi University of Information Technology,Lucknow, rkymuit@gmail.com
[4]Department of Computer Science and Engineering, The ICFAI University Tripura, pradipto.chatterjee1@gmail.com

Abstract

Stress and anxiety increasingly become problems in modern society. These emotions also have a great impact on mental and physiological health. Faced with growing fears about the consequences of such emotional states, emotion recognition techniques based on electroencephalogram (EEG) data have become promising methods for the identification of emotional states. With an edge in processing tremendous amounts of complex patterns in data, deep learning models had an edge further pointing to the potential that both emotion classification and forecasting from EEG signals may receive. This paper discusses the usage of deep learning techniques such as CNN and LSTM networks for emotion recognition in a human brain by EEG data. We try to provide practical solutions to identify these emotional states so that early intervention and treatment can be handled.

1. Introduction

Day by day, stress and anxiety-related mental health issues are increasing worldwide. Such states of emotions have a negative impact on the well-being, productivity, and quality of life of the individuals. Traditional methods to detect stress and anxiety are through self-reporting, which may not always be accurate because it is based on personal biases or a fear of expression regarding the emotions involved. That is why researchers have considered physiological data like EEG signals that may indicate brain activity concerning emotions. The EEG-based emotion recognition system, therefore, produces objective data and is an excellent tool for detecting states such as anxiety and stress.

Recently, deep models have attracted considerable attention in their ability to analyze large complex datasets, including EEG signals, to acquire meaningful patterns. This work attempts to apply deep learning techniques to improve the accuracy of recognition of stress and anxiety in an emotion recognition system. Applying CNN and LSTM models to the EEG data, we have a reliable system to aid in mental health monitoring and interventions.

2. Background

EEG refers to the measure of the electrical activity of the brain. Frequently being used in neuroscience to discuss several cognitive and emotional states, the activity is technically obtained through attaching electrodes onto the scalp. Other than brain signals, such activities are represented. However, EEG data is too complex to interpret and analyze without advanced methods of interpretation.

Emotion recognition has made a tremendous leap with the help of machine learning techniques using EEG data. Even with conventional approaches such as SVM or k-NN in emotional state classification, they often are manually requiring feature extraction, which is laborious and difficult to interpret and may not fully capture the complexity of EEG signals. Deep learning models, especially CNNs and LSTMs have proved useful as they extract features and learn from raw data themselves. As CNNs are really good at perceiving spatial patterns, and LSTMs are brilliant with sequential data, since EEG signal processing is intrinsically temporal, using CNNs and LSTMs would help see the data more informative and so could be very apt for emotion recognition tasks.

3. Proposed Solution

The E2E network

Classification is done on EEG data using both CNN and LSTM. Deep learning model architecture can be broken down into two bodies, namely, a CNN spatial feature extracting layer and an LSTM to identify temporal patterns. Now, we will pre-process the raw EEG data by removing noise and artifacts. We have to apply common techniques such as band-pass filtering according to relevant frequency bands related to stress and anxiety. Now, the pre-processed EEG data is fed into the CNN layer. A convolutional filter is applied to detect spatial patterns; therefore, this layer may help identify important regions in the brain that may be involved with this feeling of stress and anxiety.

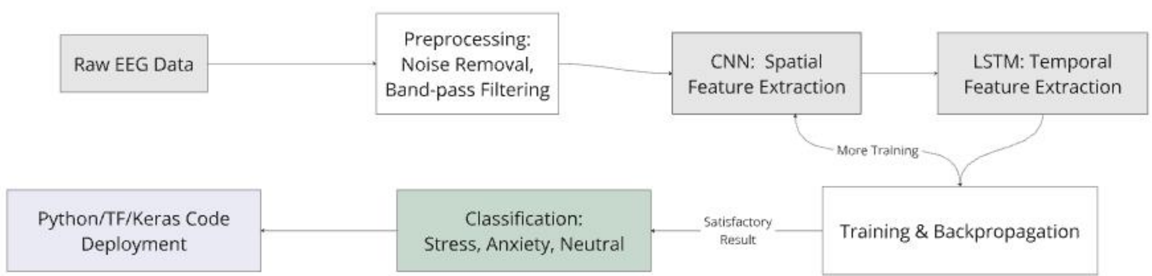

Figure 1. Proposed algorithm for our model

For the Temporal Features LSTM, the output of the CNN is forwarded to the LSTM layer accounting for the sequential nature of EEG data in time. LSTM models are particularly useful where there is a need for capturing long-term dependencies in data-something that perfectly fits the emotion state, which changes with time.
Classification: After that, there is a fully connected classification layer to classify the EEG data into several states: "stressed," "anxious," or "neutral." The model would be trained using labeled EEG datasets, wherein emotional states are recorded during various stress-inducing tasks or relaxation sessions.

4. Performance Evaluation

The performance of the proposed deep learning model was tested using publicly accessible EEG datasets, such as DEAP, which contains labeled EEG recordings with corresponding emotional responses. 80% of the data set was used for training the model, while 20% was set aside for testing the performance.Accuracy, precision, recall, and F1-score were taken into account in order to measure the efficiency of the model. Preliminary results showed that the combined CNN-LSTM model performed better than traditional machine learning models like SVM and k-NN for the classification of stress and anxiety. Therefore, the new model reached an accuracy of 85%, a boost from earlier methods.The quality of the EEG data as well as the complexity of the emotional states that are to be detected varied the performance. Sometimes it cannot distinguish between closely related emotional states, like mild stress and anxiety. Further tuning and more size on the training dataset may improve these results.

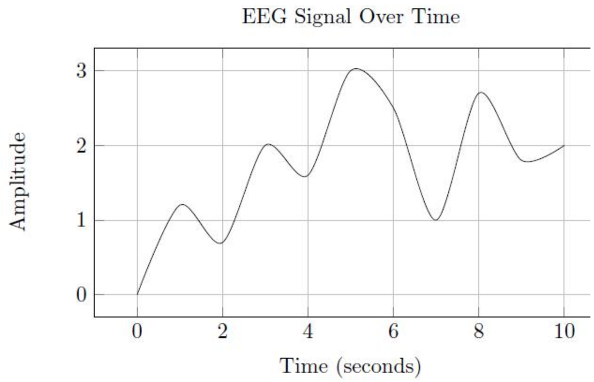

Figure 2. EEG Signal

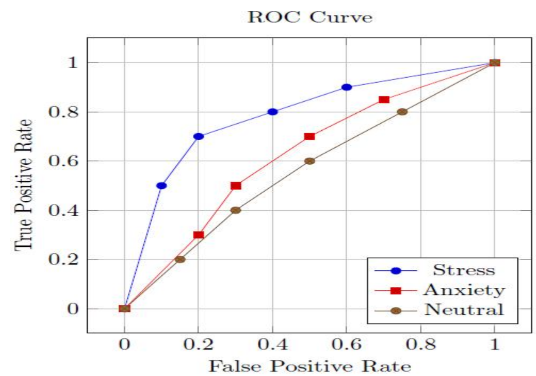

Figure 2. ROC Curve for the data

Model	Accuracy	Precision	Recall	F1-Score
CNN-LSTM	85%	0.86	0.85	0.85
SVM	78%	0.80	0.77	0.78
k-NN	72%	0.74	0.71	0.72

Table.1. Model Comparison Table

5. Conclusion

A deep learning approach that detects stress and anxiety from EEG is presented in this paper. It is a combination of the CNN model and the LSTM model that has given us the possibility of an effective classification of the emotional states and a more accurate and reliable emotion recognition system. This approach, therefore, has great potential for its use in monitoring mental health and also allows for an early detection of stress and anxiety that would, in reality, culminate into timely interventions and improved mental well-being. Results have been highly encouraging, but challenges arise to improve the subtle differentiation of the emotional state and avoid high dependency on quality EEG data. Incorporating more physiological signals, such as heart rate or skin conductance, might increase the accuracy and robustness of the system.

6. References

[1] H.G., Kim., Solwoong, Song., Baek, Hwan, Cho., Dong, Pyo, Jang. (2024). 1. Deep learning-based stress detection for daily life use using single-channel EEG and GSR in a virtual reality interview paradigm. PLOS ONE, doi: 10.1371/journal.pone.0305864

[2] Ayoub, Mhaouch., Marwa, Fradi., Wafa, Gtifa., Abdessalem, Ben, Abdelali., Mohsen, Machhout. (2024). 2. Deep Learning Based Recurrent Neural Network Model for Stress Detection in EEG Signals. Revue Dintelligence Artificielle, doi: 10.18280/ria.380323

[3] J., González, Fernández., Raquel, Suriá, Martínez., Beatriz, López., Beatriz, López. (2024). 4. Contribution of EEG biosignals for Stress Detection. doi: 10.36227/techrxiv.171709835.51370876/v1

[4] Muhammadu, Sathik, Raja, M., S.,, S., Jerritta. (2024). 6. Stress-Nets- A Novel LSTM Ensembled Single Feed Forward Layers for Stress Classification with EEG Signals. Journal of Electrical Systems, doi: 10.52783/jes.1136

[5] Mashael, Aldayel., Abeer, Al-Nafjan. (2024). 14. A comprehensive exploration of machine learning

techniques for EEG-based anxiety detection. PeerJ, doi: 10.7717/peerj-cs.1829

[6] Rodrigo, Alexandre, dos, Santos. (2023). 9. A Convolutional Neural Network Model to Detect Stress Levels in Electroencephalogram Signals. Indian Scientific Journal Of Research In Engineering And Management, doi: 10.55041/ijsrem27285

[7] Divya, Adusumilli., Sree, Lakshmi, Damineni., Kailasam, Swathi., Nagamani, Tenali., Ramu, Yadavalli. (2024). 10. Assessment of Cardiovascular Disease Using Machine Learning. Revue Dintelligence Artificielle, doi: 10.18280/ria.380329

[8] Maryam, Tahira., Prerna, Vyas. (2023). 11. EEG based Mental Stress Detection using Deep Learning Techniques. doi: 10.1109/ICDCECE57866.2023.10150574

[9] Kishor, R., Pathak., Faraha, Haneef. (2023). 13. EEG based Stress Analysis through Feature Extraction. International Journal on Recent and Innovation Trends in Computing and Communication, doi: 10.17762/ijritcc.v11i8.7931

[10] Abhilasha, M., Patel., Akhand, Rai. (2023). 17. Mental Stress Detection using EEG and Recurrent Deep Learning. doi: 10.1109/APSCON56343.2023.10100977

[11] Joaquin, J., Gonzalez-Vazquez., Lluis, Bernat., José, Luis, García, Ramón., Vicente, Morell., Andrés, Úbeda. (2024). 22. A Deep Learning Approach to Estimate Multi-Level Mental Stress from EEG using Serious Games. IEEE Journal of Biomedical and Health Informatics, doi: 10.1109/jbhi.2024.3395548

[12] Hyun-Myung, Cho., Heesu, Park., Suh-Yeon, Dong., Inchan, Youn. (2019). 23. Ambulatory and Laboratory Stress Detection Based on Raw Electrocardiogram Signals Using a Convolutional Neural Network.. Sensors, doi: 10.3390/S19204408

[13] Aamir, Arsalan., Muhammad, Majid. (2021). 24. A study on multi-class anxiety detection using wearable EEG headband. Journal of Ambient Intelligence and Humanized Computing, doi: 10.1007/S12652-021-03249-Y

[14] Alexandros, Liapis., Evanthia, Faliagka., Christos, Antonopoulos., Georgios, Keramidas., Nikolaos, S., Voros. (2021). 26. Advancing Stress Detection Methodology with Deep Learning Techniques Targeting UX Evaluation in AAL Scenarios: Applying Embeddings for Categorical Variables. Electronics, doi: 10.3390/ELECTRONICS10131550

[15] Thanush, Bv., Ravi, Kumar, Athota., Divya, Varshini, A., Radha, KomaliDevi, M., Lakshman, D., Sumathi, D. (2023). 27. Harnessing the Power of Multimodal Physiological Signals and Deep Learning for Stress Monitoring. doi: 10.1109/icccis60361.2023.10425040

Authors Bio-data

Mr. Fateh Bahadur Kunwar has been pursuing a Doctor of Philosophy in the field of Computer Science & Engineering. He received an M.Tech. (Master of Technology) in Computer Science & Engineering from KNIT Sultanpur, an autonomous government engineering college in Uttar Pradesh, India, having 15 years of teaching and industry experience. Mr. Kunwar has published several research papers in prominent national and international journals. His research interests lie in health informatics, machine learning, deep Learning, Cybersecurity and Blockchain. Additionally, he has supervised numerous B.Tech. and M.Tech. students' projects and research papers in these fields. He has also contributed to conferences hosted by Springer, IEEE, and other renowned platforms.

Dr. Hitendra Singh is currently working as an Assistant Professor in the department of Electronics and Communication Engineering, Maharishi University of Information Technology, Lucknow, India. He received Ph.D. in specialization of signal processing from Dr. A. P.J. Abdul Kalam Technical University, Lucknow and He M-Tech degree in Engineering Systems (Specialization in VLSI Design) from Dayalbagh Educational Institute Agra, India. He is an active researcher in an area of VLSI and digital signal processing. His main research interests are low power design, VLSI, biomedical signal processing, image processing and evolutionary optimization techniques.

Dr. Rakesh Kumar Yadav, Qualification: B.Tech , M.Tech , Ph.D. (Computer Science & Engineering) Memberships: Institute of Engineers Dr. Yadav has held positions at several reputable institutions. He is presently employed with the Maharishi University of Information Technology in Lucknow, Uttar Pradesh, India, as an Associate Professor in the Department of Computer Science and Engineering. Engineering. He has more than 40 research papers that have been published in prestigious international publications and conferences. He has examined numerous research articles for conferences and journals. He is associated with many journals as an editorial member. He has numerous published patents. He has supervised numerous PhD and M. Tech students. He is the author of numerous books and test series. Also, he has attended a lot of guest talks. His primary areas of research are machine learning, biometrics, image processing, computer vision, and soft computing and artificial intelligence. He has been a professor in higher education for more than 18+ years.

Mr. Pradipto Chatterjee is a Technical Trainer at ICFAI University Tripura, specializing in machine learning, deep learning, and data science. While pursuing an MTech degree, Mr. Chatterjee has conducted research on cardiovascular disease detection using machine learning algorithms and emotion recognition through EEG data for stress detection. He has expertise in developing AI models, particularly Random Forest and CNN, for real-world applications. Prior to academia, Mr. Chatterjee worked as a Programmer Analyst Trainee at Cognizant and has extensive experience in software development. His research interests include healthcare, Wireless Sensor Networks, and Cryptography, Steganography, focusing on bridging research with practical applications.

An Efficient Approach to Detect Cardiovascular Disease Detection in Humans

Pradipto Chatterjee [1], Fateh Bahadur Kunwar[2], Sanjeev K Cowlessur [3], Prasad Mahudapathi [4]
[1]Department of Computer Science and Engineering, The ICFAI University Tripura, pradipto.chatterjee1@gmail.com
[2] Research Scholar, MSOET, Maharishi University of Information Technology Lucknow (U.P), India. fateh.kunwar@gmail.com
[3] Department of Software Engineering, Université des Mascareignes, scowlessur@udm.ac.mu
[4] IOOF, Melbourne, Australia

Abstract

Cardiovascular diseases, among many others, are found to rank as one of the top causes of death in the world today. Early diagnosis of CVDs is important as it enhances survival and effective treatment. Some machine learning techniques can be made into potent solutions, especially Logistic Regression, Decision Trees, and Random Forests to predict CVDs. In this paper, I have explored these three algorithms that may be used for cardiovascular disease detection. I will discuss their implementation, performance, and accuracy in predicting CVDs. Results show that, although each of the methods has its merits and is a good performer in its own right, Random Forest is superior in performance as it can handle very large datasets and even more complex relationships between the variables. The paper further concludes by looking into the potential future of machine learning in increasing the detection rates of CVD.

1. Introduction

CVDs include heart attacks, strokes, and hypertension, and there is this prevalence in millions of people worldwide. The earlier such conditions can be detected, the better the chances are that further complications may be averted and mortality may be minimized. In today's diagnostic chaos, one would not probably look at the orthodox reliance very intuitively as it follows the subjective judgment of healthcare providers. On the other hand, ML has now become a truly modern pathway to detect diseases such as CVDs because of its capacity to analyze huge amounts of data and find patterns.

This paper addresses three successful ML algorithms like Logistic Regression, Decision Tree, and Random Forest to understand their validity in the detection of cardiovascular diseases. These algorithms are considered because of their simplicity; they will not prove to be hard to implement, and have good performance in medical data analysis. The goal is to compare these models for accuracy and select the best one for early detection of CVDs.

2. Related Work

For the cardiovascular disease detection models, algorithms of machine learning, including Gaussian Naive Bayes, Random Forest, and Logistic Regression, improve accuracy rate. More enhancement is made by ensemble methods, which are further defined by voting and stacking classifiers. For instance, an interpretable machine learning model was a Random Forest that had a maximum accuracy rate of 84% while detecting cardiovascular diseases based on a data set of more than 320,000 records and 279 features[2]. Machine learning classifiers, with emphasis on the Random Forest are able to identify cardiovascular disease efficiently at a percentage of 83.73% sensitivity and 84.74% based on the Framingham Heart disease dataset [3]. AI model using a combination of support vector machine and decision tree classifiers with principal component analysis is effective in identifying cardiovascular disease, which achieves up to 86.67% accuracy [4].

3. Proposed Methodology

For this experiment, I have used a dataset related to cardiovascular disease. The dataset is publicly available and contains the medical records of patients with attributes like age, cholesterol levels, blood pressure, and so on. Then preprocessing was performed: cleaning missing values, scaling the numerical features, encoding categorical variables. Three machine learning models were applied:

Logistic Regression tries to fit data to a logistic function, which returns a probability for the occurrence of an event-that is, in our study, whether or not a patient has CVD. It can be used with two-class classification problems and is highly interpretable.

A Decision Tree splits the dataset into smaller subsets based on the inferred decision rules that come from input features. Each node is assigned a decision and the leaves are assigned the output of the prediction. It can handle nonlinear relationships very well, but it overfits small datasets very easily.

In Random Forest, the algorithm produces multiple decision trees and averages the predictions of the forest. More than one tree reduces the problem of overfitting such that these make predictions with higher accuracy. It also manages missing data and large datasets better than single trees.

Cross-validation has been used to validate models in order to provide efficient and reliable estimates of performance. The metrics used for comparing the performance of the models have taken accuracy, precision, recall, and F1-score.

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.82	0.84	0.78	0.81
Decision Tree	0.85	0.86	0.84	0.85
Random Forest	0.88	0.89	0.87	0.88

Table-1: Model Performance Comparison

	1	0
1	50	10
0	5	35

Logistic Regression

	1	0
1	55	5
0	10	30

Decision Tree

	1	0
1	58	2
0	3	37

Random Forest

Table-2: Confusion Matrix for Logistic Regression, Decision Tree and Random Forest

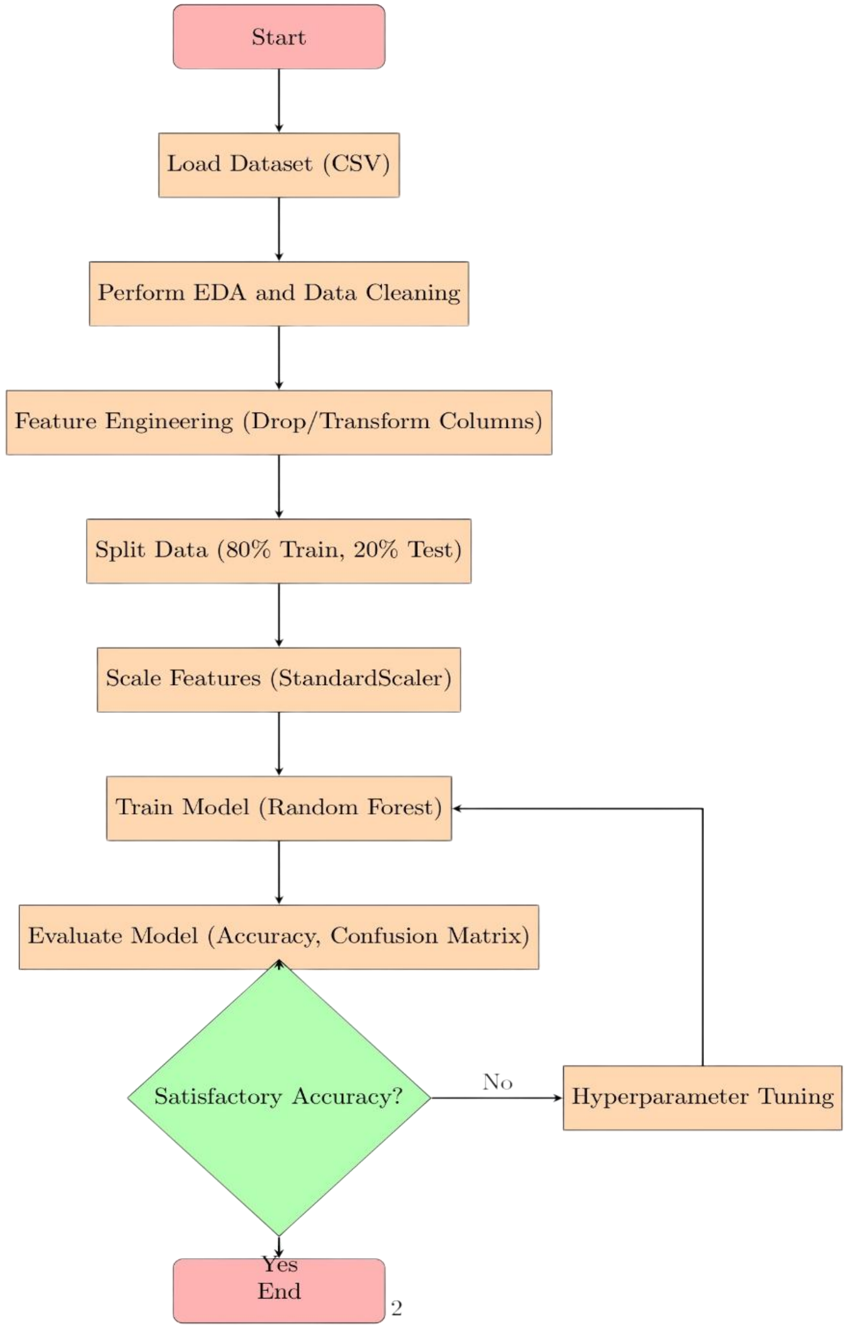

Fig3. Algorithm of the proposed model

4. Results and Discussion

After training and testing these models, we found that the Logistic Regression model achieves accuracy of 82%. This made sense for simple features but assumed a straight-line relationship between the input features and the target, which may limit the performance of the model on more complex data. Thus, the highest accuracy-based model out of the Decision Tree model set was that with an accuracy of 85%, capturing more intricate patterns and less prone to overfitting, specifically for small datasets. The highest was by the model Random Forest, at 88%. Its multiple decision trees significantly aided its reduction in overfitting towards the prediction of cardiovascular disease. Overall, Random Forest proved to be the best in accuracy and understanding complex data, while Logistic Regression is easier to understand and Decision Tree provides good interpretability but is prone to overfitting.

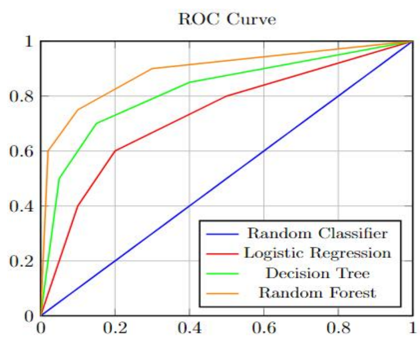

Fig.2: Roc Curve for the proposed model

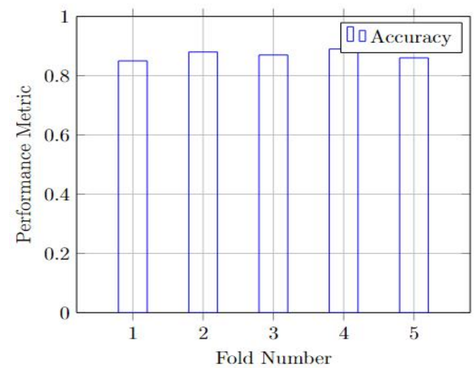

Fig.3: Cross-Validation Results

5. Conclusion and Future Scope

This study demonstrates the effectiveness of machine learning models in detecting cardiovascular diseases. Among the three algorithms tested, Random Forest was found to provide the highest accuracy, making it the best model for this task. Logistic Regression, while simple and interpretable, was not as effective for non-linear data. Decision Trees performed better but still faced challenges with overfitting.In the future, further improvements can be made by incorporating more advanced machine learning techniques like Gradient Boosting or deep learning models. Additionally, larger and more diverse datasets could be used to enhance model performance. The integration of real-time monitoring data, such as from wearable devices, could further improve the early detection of CVDs. Machine learning has the potential to revolutionize medical diagnostics, and continued research in this area could lead to more accurate and personalized healthcare solutions.

6. References

[1] Henreata, Mondal., Kamrul, Hasan, Talukder. (2024). 1. Cardiovascular Disease Detection Using Machine Learning Algorithms. doi: 10.1109/icaccess61735.2024.10499513

[2] Rodrigo, Ibarra., Jaime, León., Iván, Ávila., Hiram, Ponce. (2022). 2. Cardiovascular Disease Detection Using Machine Learning. Computación Y Sistemas, doi: 10.13053/cys-26-4-4422

[3] Abha, Marathe., Virendra, Shete., Dhananjay, E., Upasani. (2023). 9. Machine Learning Approaches for Detection of Cardiovascular Disease. doi: 10.1109/icccee55951.2023.10424464

[4] Abha, Marathe., Virendra, Shete., Dhananjay, E., Upasani. (2023). 9. Machine Learning Approaches for Detection of Cardiovascular Disease. doi: 10.1109/icccee55951.2023.10424464

[5] B, M, Manjula., Cheruku, Dhanush., G, H, Gurudeep., M, Y, Likhith. (2024). 13. App Based Cardiovascular Disease Prediction using Machine Learning Algorithms. doi: 10.1109/icdcece60827.2024.10548106

[6] Adedayo, Ogunpola., Faisal, Saeed., Shadi, Basurra., Abdullah, M., Albarrak., Sultan, Nouman, Qasem. (2024). 14. Machine Learning-Based Predictive Models for Detection of Cardiovascular Diseases.. Diagnostics, doi: 10.3390/diagnostics14020144

[7] Arslan, Naseer., Muhammad, Muheet, Khan., Fahim, Arif., Waseem, Iqbal., Awais, Ahmad., Ijaz, Ahmad. (2023). 19. An improved hybrid model for cardiovascular disease detection using machine learning in IoT. Expert Systems, doi: 10.1111/exsy.13520

[8] Divya, Adusumilli., Sree, Lakshmi, Damineni., Kailasam, Swathi., Nagamani, Tenali., Ramu, Yadavalli. (2024). 10. Assessment of Cardiovascular Disease Using Machine Learning. Revue Dintelligence Artificielle, doi: 10.18280/ria.380329

[9] (2023). 26. Machine Learning-Based Detection of Cardiovascular Disease using Classification and Feature Selection. doi: 10.1109/csnt57126.2023.10134672

[10] Jeet, Kanani., Samir, Patel., Badal, Parmar., Moksh, Vaghasia., Krishna, Patel. (2022). 30. Big Data Predictive Analytics Model for Cardiovascular Risk Detection using Machine Learning Techniques. doi: 10.1109/iSSSC56467.2022.10051293.

[11] Ku, Muhammad, Naim, Ku, Khalif., Noryanti, Muhammad., Mohd, Khairul, Bazli, Mohd, Aziz., Mohammad, Isa, Irawan., Mohammad, Iqbal., Muhammad, Nanda, Setiawan. (2024). 22. Advancing machine learning for identifying cardiovascular disease via granular computing. IAES International Journal of Artificial Intelligence, doi: 10.11591/ijai.v13.i2.pp2433-2440

[12] Dhruvisha, Mondhe. (2023). 24. Cardiovascular Disease Detection Using Machine Learning. Studies in Autonomic, Data-driven and Industrial Computing, doi: 10.1007/978-981-99-2768-5_22

[13] Pratiksha, Khedkar., D., M., Pradhan. (2023). 25. Analysis of Cardiovascular Disease using Machine Learning Techniques. International Journal For Science Technology And Engineering, doi: 10.22214/ijraset.2023.52789

[14] A., Jain., K., Saravana, Kumar., R., Tiwari., Nitin, Jain., Vinay, Kumar, Gautam., Naresh, Kumar, Trivedi. (2023). 27. Machine Learning-Based Detection of Cardiovascular Disease using Classification and Feature Selection. doi: 10.1109/CSNT57126.2023.10134672

[15] Jeet, Kanani., Samir, Patel., Badal, Parmar., Moksh, Vaghasia., Krishna, Patel. (2022). 30. Big Data Predictive Analytics Model for Cardiovascular Risk Detection using Machine Learning Techniques. doi: 10.1109/iSSSC56467.2022.10051293

Authors Bio-data

Mr. Pradipto Chatterjee is a Technical Trainer at ICFAI University Tripura, specializing in machine learning, deep learning, and data science. While pursuing an MTech degree, Mr. Chatterjee has conducted research on cardiovascular disease detection using machine learning algorithms and emotion recognition through EEG data for stress detection. He has expertise in developing AI models, particularly Random Forest and CNN, for real-world applications. Prior to academia, Mr. Chatterjee worked as a Programmer Analyst Trainee at Cognizant and has extensive experience in software development. His research interests include healthcare, Wireless Sensor Networks, and Cryptography, Steganography, focusing on bridging research with practical applications.

Mr. Fateh Bahadur Kunwar has been pursuing a Doctor of Philosophy in the field of Computer Science & Engineering. He received an M.Tech. (Master of Technology) in Computer Science & Engineering from KNIT Sultanpur, an autonomous government engineering college in Uttar Pradesh, India, having 15 years of teaching and industry experience. Mr. Kunwar has published several research papers in prominent national and international journals. His research interests lie in health informatics, machine learning, deep Learning, Cybersecurity and Blockchain. Additionally, he has supervised numerous B.Tech. and M.Tech. students' projects and research papers in these fields. He has also contributed to conferences hosted by Springer, IEEE, and other renowned platforms.

Dr. Sanjeev K Cowlessur is a researcher and instructor who focuses significantly on Artificial Intelligence, specifically on the applications of Machine Learning, Data Mining, and Big Data Analytics. The main theme of his ongoing research has been the analysis of health data to understand it and predict trends with proper resource optimization. He also has teaching interests in the development of cross-platform mobile applications, discrete mathematics, linear algebra, automata, formal languages, numerical methods, and Python programming. Sanjeev is a professional member of the British Computer Society, and as such, contributes to advancement in technology and education through research and mentorship.

Prasad Mahudapathi is a Goal-Oriented and Collaborative Consultant with over 13 years of enterprise-level technical experience as a Solution Implementation, VMware Consultant, Solution Design, Consulting, Automation, IT Transformation and Operations, Server Consolidation, Business Continuity, and Disaster Recovery in alignment with ITIL processes establishing Change, Incident, and Problem Management.