International Journal of Trend in Scientific Research and Development (IJTSRD)
Special Issue on Emerging Trends and Innovations in Web-Based Applications and Technologies
Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Advancing Horizons in Chronic Diseases: Research Innovation Insights

Vaishnavi Watkar[1], Shivani Ayyagari[2], Prof. Anupam Chaube[3]

[1,2,3]Department of Science and Technology,
[1,2,3]G H Raisoni College of Engineering and Management, Nagpur, Maharashtra, India

ABSTRACT

Technological development, including machine learning, has a huge impact on health through an effective analysis of various chronic diseases for more accurate diagnosis and successful treatment. In the field of biomedical and healthcare communities the accurate prediction plays the major role to find out the risk of the disease in the patient. The only way to overcome with the mortality due to chronic diseases is to predict it earlier so that the disease prevention can be done. Such model is a Patient's need in which Machine Learning is highly recommendable. But the precise prediction on the basis of symptoms becomes too difficult for doctor. The correct prediction of disease is the most stretching task. To overcome this problem data mining plays an important role to predict the disease. We use Heart disease, Kidney disease, Cancer disease and Diabetes disease datasets, In order to build reliable prediction models for these chronic diseases using data mining techniques. The most relevant features are selected from the dataset for improved accuracy and reduced training time. The system analyzes the symptoms provided by the user as input and gives the probability of the disease as an output Disease by using, random forest and decision tree we are predicting diseases like Diabetes, Heart, Cancer and Kidney. For each chronic disease, diverse models, techniques, and algorithms are used for predicting and analyzing. The common prediction objective is to minimize the prediction error as low as possible. The final discussions of this paper are works in improving the prediction performance for chronic diseases using a data preprocessing handling.

KEYWORDS: *Chronic Diseases, Machine Learning, Diseases Prediction, Accuracy, Prediction performance*

I. INTRODUCTION

"**Advancing Horizons in Chronic Diseases: Research Innovation Insights** "refers to discovering new opportunities and solutions for managing long-term illnesses such as diabetes, heart disease, and cancer. It emphasizes how advanced research methods and modern technologies, like AI and data analysis, are transforming how these diseases are diagnosed, treated, and prevented. The focus is on exploring the latest developments and sharing knowledge to improve healthcare outcomes. Nowadays, humans face various diseases due to the current environmental condition and their living habits.

The identification and prediction of such diseases at their earlier stages are much important, so as to prevent the extremity of it. It is difficult for doctors to manually identify the diseases accurately most of the time. The goal of is to identify and predict the patients with more common chronic illnesses. This could be achieved by using a cutting-edge machine learning technique to ensure that this categorization reliably identifies persons with chronic diseases. The prediction of diseases is also a challenging task. Hence, data mining plays a critical role in disease prediction.

With ML models, it can also be possible to improve quality of medical data, reduce variation in patient rates, and save in medical costs.

Machine learning examine the study and construction of algorithms that can learn from and make predictions on data. It is closely related to (and often overlaps with) computational statistics, which also focuses on prediction-making through the use of computers.

The preprocessing handling we discuss includes missing values, outliers, feature selection, normalization, and imbalance. The final discussions of this paper are open issues, and the potential future works in improving the prediction performance for chronic diseases using a data preprocessing handling and machine learning methods.

Our project mainly focusses on the following mentioned outcomes:-

1. Early Detection of Chronic Diseases
2. Personalized Treatment Plans
3. Enhanced Medical Data Quality
4. Smarter Health Decision-Making
5. Customizing Health Plans for Individuals

By using machine learning and deep learning algorithms this particular research paper manages to ensure chronic disease management in order to improve predictions and also advancing overall healthcare efficiency, ultimately benefiting both patients and healthcare systems.

II. RELATED WORK

This section describes the related works that are performed in developing the proposed model for predicting chronic diseases. The following are the discussions made by reviewing the existing literature that helps develop the proposed system efficiently and effectively.

Different machine-learning techniques have been used for effective classification of chronic kidney disease from patients' data which are described as follows:-

Charleonnan et al. [8] did comparison of the predictive models such as K-nearest neighbors (KNN), support vector machine (SVM), logistic regression (LR), and decision tree (DT) on Indians Chronic Kidney Disease (CKD) dataset in order to select best classifier for predicting chronic kidney disease. Tey have identified that SVM has the highest classification accuracy of 98.3% and highest sensitivity of 0.99

Priyanka et al. [12] carried out chronic kidney disease prediction through naive bayes. Tey have tested using other algorithms such as KNN (K-Nearest Neighbor Algorithm), SVM (Support Vector Machines), Decision tree, and ANN (Artificial Neural Network) and they have got Naïve Bayes with better accuracy of 94.6% when compared to other algorithms.

Mohammed and Beshah [13] conducted their research on developing a self-learning knowledge-based system for diagnosis and treatment of the first three stages of chronic kidney have been conducted using machine learning. A small number of data have been used in this research and they have developed prototype which enables the patient to query KBS to see the delivery of advice. Tey used decision tree in order to generate the rules. The overall performance of the prototype has been stated as 91% accurate.

Salekin and Stankovic [9] did evaluation of classifiers such as K-NN, RF and ANN on a dataset of 400.

Wrapper feature selection were implemented and five features were selected for model construction in the study. The highest classification accuracy is 98% by RF and a RMSE of 0.11. S. Tekale et al. [10] worked on "Prediction of Chronic Kidney Disease Using Machine Learning Algorithm" with a dataset consists of 400 instances and 14 features. Tey have used decision tree and support vector machine. Te dataset has been preprocessed and the number of features has been reduced from 25 to 14. SVM is stated as a better model with an accuracy of 96.75%.

Xiao et al. [11] proposed prediction of chronic kidney disease progression using logistic regression, Elastic Net, lasso regression, ridge regression, support vector machine, random forest, XGBoost, neural network and k-nearest neighbor and compared the models based on their performance. Tey have used 551 patients' history data with proteinuria with 18 features and classified the outcome as mild, moderate, Debal and Sitote Journal of Big Data (2022) 9:109 Page 3 of 19 severe. Tey have concluded that Logistic regression performed better with AUC of 0.873, sensitivity and specificity of 0.83 and 0.82, respectively.

Almasoud and Ward [13] aimed in their work to test the ability of machine learning algorithms for the prediction of chronic kidney disease using subset of features. Tey used Pearson correlation, ANOVA, and Cramer's V test to select predictive features. Tey have done modeling using LR, SVM, RF, and GB machine learning algorithms. Finally, they concluded that Gradient Boosting has the highest accuracy with an F-measure of 99.1.

Most previously conducted researches focused on two classes, which make treatment recommendations difficult because the type of treatment to be given is based on the stages as our project focuses on chronic disease prediction using machine learning models based on the dataset with big size and recent than online available dataset

III. PROPOSED WORK

Due to the low-progress nature of Chronic Diseases, it is important to make an early prediction and provide effective medication. Therefore, it is essential to propose a decision model which can help to diagnose chronic diseases and predict future patient outcomes. While there are many ways to approach this in the field of AI, the present study focuses distinctly on ML predictive models used in the diagnosis of Chronic Diseases. In comparison to the conventional data analysis techniques, we will be able to find promising results that enhance the quality of patient data and inspect of specific items that are related to ML algorithms in medical care.

The main purpose of our research paper is to make hospital tasks easy and to develop an efficient and feasible software that replaces the manual prediction system into an automated healthcare management system and also it enables healthcare providers to improve operational effectiveness, reduce medical errors and time consumption. If disease can be predicted, then early treatment can be given to the patients which can reduce the risk of life and save life of patients. The cost to get treatment of diseases can also be reduced up to an extent by early recognition.

Our proposed framework aims to predict chronic diseases such as heart, kidney, cancer, and diabetes using:

- **A Hybrid Architecture**: Combines features from text data (processed with TF-IDF) and medical images (processed using ResNet-18).
- **Multi-Modal Learning**: Merges textual and visual modalities to improve prediction accuracy compared to single-modal systems..
- **Scalability**: Designed to handle additional data types, ensuring broad applicability in clinical settings.

Table 1: Data Preprocessing Steps

Step	Text Data	Image Data
Feature Extraction	TF-IDF Transformation	ResNet-18 Feature Extraction
Normalization	Numeric Standardization	RGB Normalization
Encoding	Label Encoding, Binary Encoding	N/A
Feature Dimensions	1200	128

IV. PROPOSED RESEARCH MODEL

The proposed model integrates both machine learning (ML) and deep learning (DL) techniques to predict chronic diseases. It adopts a hybrid multi-modal approach, combining text-based medical data analysis with image-based diagnostics to enhance prediction accuracy.

1. Text-Based Analysis

- **Feature Extraction**: Patient symptoms, lifestyle, and medication data are transformed using TF-IDF vectorization and categorical encoding.
- **Model Input**: Text features (e.g., symptoms, demographic details) with up to 1200 dimensions.

2. Image-Based Analysis

- **Feature Extraction**: Medical images (X-rays, CT scans) are processed through a pre-trained ResNet-18 model for feature extraction (128 dimensions).
- **Preprocessing Steps**: RGB conversion, resolution standardization, normalization, and augmentation for improved model robustness.

3. Hybrid Integration

- Text and image features are combined into a unified vector of 1328 dimensions.
- A neural network processes this combined feature vector using two hidden layers with dropout for regularization, culminating in disease classification.

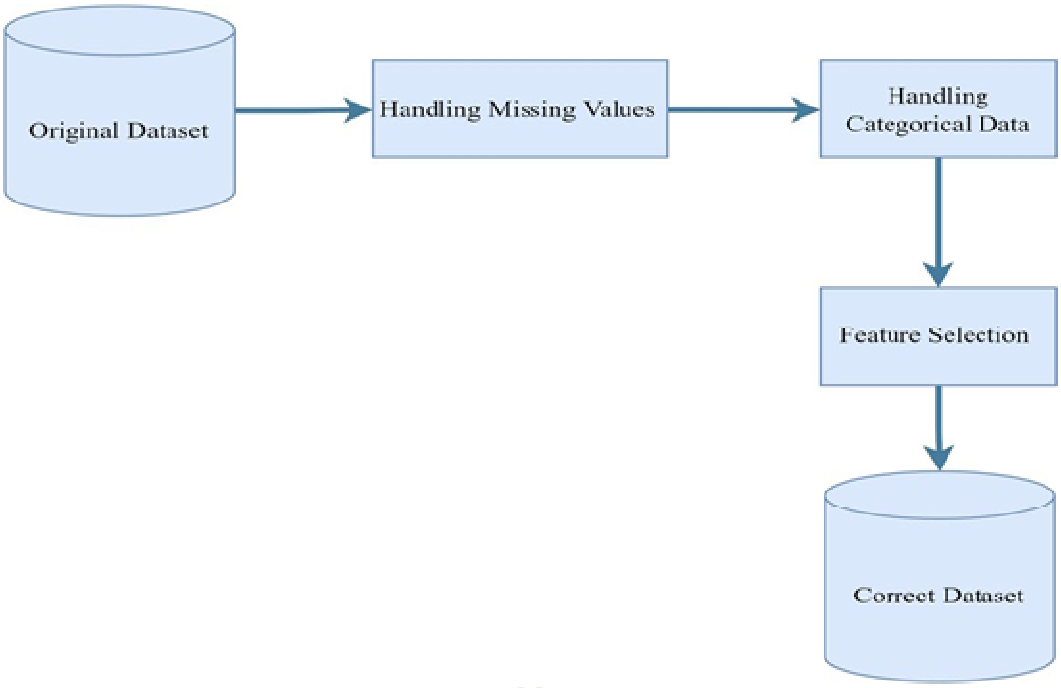

Fig 1: Chronic disease dataset processing steps

➢ **Original Dataset :**

- This is the starting point, containing raw data in its original form. It might include missing values, categorical features, and irrelevant features.

➢ **Handling Missing Values :**

- Missing data can significantly impact the accuracy of machine learning models. This step involves techniques like:
- Deletion: Removing rows or columns with missing values.
- Imputation: Replacing missing values with estimated values (e.g., mean, median, mode, or more sophisticated methods).

➢ **Handling Categorical Data :**

- Many machine learning algorithms require numerical data. This step converts categorical features (e.g., "male", "female", "red", "blue") into numerical representations:
- One-Hot Encoding: Creating binary columns for each category (e.g., "male" becomes "male_1", "female" becomes "female_1").
- Label Encoding: Assigning a unique integer to each category.
- Other techniques: Depending on the data and algorithm.

➢ **Feature Selection :**

- This step aims to identify and select the most relevant features for the machine learning model. This can improve model performance and reduce training time:
- Filter methods: Evaluating features based on their individual scores (e.g., correlation with target variable).
- Wrapper methods: Selecting features based on their performance in a model (e.g., recursive feature elimination).
- Embedded methods: Integrating feature selection into the model training process (e.g., L1 regularization).

➢ **Correct Dataset :**

- The final output is a preprocessed dataset ready for machine learning algorithms. It has no missing values, categorical features are encoded, and only the most relevant features are retained.

These steps proves to be helpful in order to carry out the main purpose of our research paper as the mentioned steps plays an important role in defining the basic structure of how data is being preprocessed in such a way that it can be able to identify or predict a particular chronic disease with the help of using machine learning and deep learning algorithms which makes any type of complex data easy to understand and interpret by the computer in order to carry out further processing of our project.

The diagram given below highlights the importance of data preprocessing and splitting data into training and testing sets for building robust machine learning models. It showcases two different approaches (CNN and KNN) that can be used for disease prediction based on the nature of the data and the desired level of complexity. The goal is to develop a predictive model that can accurately diagnose diseases and assist healthcare professionals in decision-making.

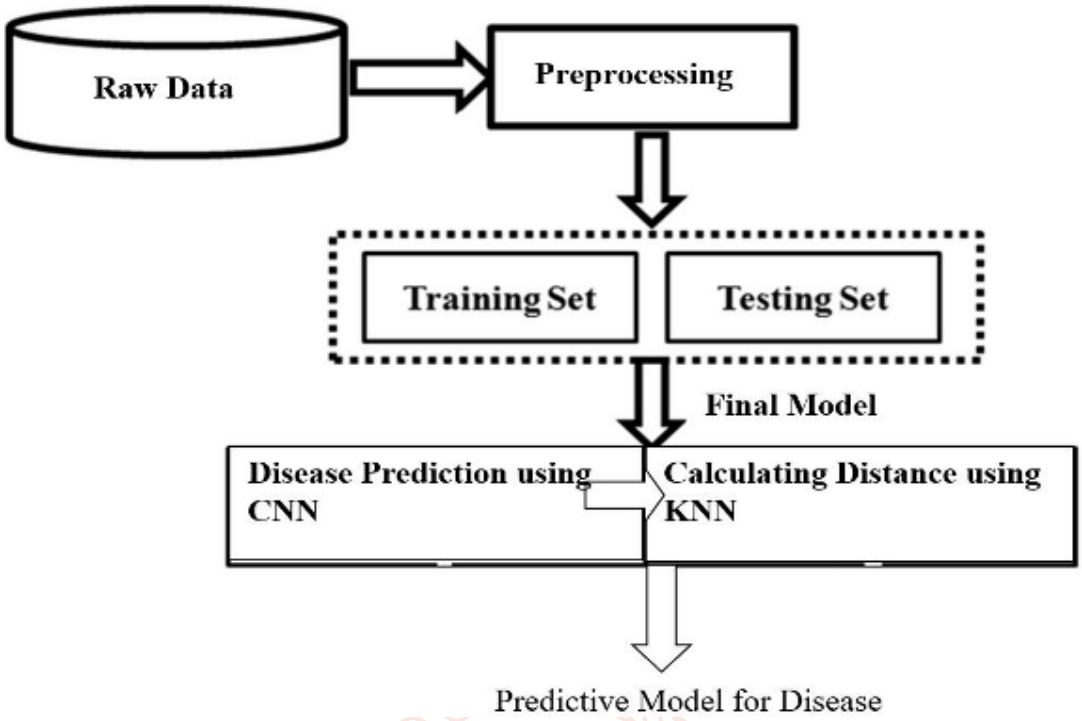

Fig 2: Proposed model of the disease predicting system

➢ **Raw Data:**

- This is the starting point, containing unprocessed information about patients and their health conditions.

➢ **Preprocessing :**

- The raw data is transformed into a suitable format for machine learning algorithms. This involves tasks like:
  - Cleaning the data to handle missing values or inconsistencies.
  - Feature engineering to extract relevant information from the raw data.
  - Normalization or scaling of data to ensure features have comparable ranges.

➢ **Training Set & Testing Set:**

- The preprocessed data is divided into two subsets:
  - Training Set: Used to train the machine learning model. The model learns patterns and relationships from this data.
  - Testing Set: Used to evaluate the model's performance on unseen data. This helps assess how well the model generalizes to new cases**.**

➢ **Final Model:**

- The diagram suggests two different approaches to building the final model **:**
  - Disease Prediction using CNN: Convolutional Neural Networks (CNNs) are deep learning models well-suited for image analysis. They might be used to analyze medical images (like X-rays or scans) to predict the presence of a disease.
  - Calculating Distance using KNN: K-Nearest Neighbors (KNN) is a simpler algorithm that classifies new data points based on the distances to their nearest neighbors in the training set.

➢ **Predictive Model for Disease:**

- The final output is a trained model that can be used to predict the presence or absence of a disease in new patients based on their data.

V. PERFORMANCE EVALUATION

The performance of the proposed model will be evaluated using the following metrics:

1. **Accuracy**: Measures the percentage of correct predictions.
2. **Class-Wise Performance**: Assesses precision, recall, and F1-score for each disease.
3. **Confidence Scores**: Analyzes the reliability of model predictions.
4. **Comparative Analysis**: Compares the hybrid model's performance to single-modal ML or DL systems.
5. **Scalability and Resource Efficiency**: Evaluates the computational efficiency and adaptability of the model to diverse datasets.

This evaluation ensures that the model provides not only accurate predictions but also scalable and resource-efficient solutions for real-world applications.

VI. RESULT ANALYSIS

➢ **Metrics Analysis :**

- Prediction Accuracy**:** Achieved high overall accuracy, demonstrating reliable performance in predicting Chronic diseases.
- Confidence Scores: The model provided consistent confidence levels, indicating robust predictions.
- Probability Distribution: Predicted probabilities across diseases were balanced, reducing bias toward specific diagnoses.
- Per-Class Metrics: Precision, recall, and F1-scores highlighted strong performance across all disease categories, visualized using a confusion matrix**.**

➢ **Comparative Analysis :**

- Better Than Single-Modality Systems: The multi-modal approach performed better than single-modality models, improving accuracy and diagnosis precision.

- Resource Usage: The model required reasonable resources, making it efficient and practical.
- Scalability: It handled larger datasets well, maintaining good performance as data increased.

➢ **Key Insights :**

- Results were visualized using graphs like ROC curves and confusion matrices.
- The system proved effective and scalable, offering reliable predictions for chronic diseases.

VII. CONCLUSION

Machine learning has made healthcare better by making it easier and more reliable to diagnose serious diseases like heart, kidney, cancer, and diabetes. Our study achieved about 90% accuracy in predicting these diseases and provides reports showing the chances of having a disease. This shows that our approach is effective and useful. The proposed model also generates detailed reports highlighting the likelihood of disease occurrence, showcasing the reliability and effectiveness of this approach.

This research paper aims to create a robust and efficient model for predicting chronic diseases using a combination of machine learning and deep learning techniques. By leveraging both text and image data, the model will provide comprehensive insights into patient health, facilitating early intervention and improved outcomes.

VIII. FUTURE SCOPE

In the future, researchers can try different types of machine learning methods, like supervised and unsupervised techniques, to see which ones work best for predicting diseases. This particular research paper can help find models that are even more accurate and reliable. Additionally, using larger datasets that include more variety—such as data from people of different ages, regions, and health conditions—can make the model work better for all kinds of patients. By also focusing on new ways to measure the model's performance, like checking how well it works in real-life situations, predictions can become more trustworthy and useful for doctors and patients.

REFERENCES

[1] Usha Kaushik Kulkarni1, Manjunath B2, Mayur Hebbar T M3, Meghana M4, Shashank S5, and Tojo Mathew6 "Chronic Disease Prediction Using Machine Learning" Vol. 10, Issue 6, June 2021 DOI 10.17148/IJARCCE.2021.10663

[2] Kang Adiwijaya, Nur Ghaniaviyanto Ramadhan, Warih Maharani and Alfian Akbar Gozali "Chronic Diseases Prediction Using Machine Learning With Data Preprocessing Handling: A Critical review" IEEE Access PP(99):1-1: January 2024 DOI:10.1109/ACCESS.2024.3406748

[3] Rakibul Islam, Azrin Sultana and Mohammad Rashedul Islam "A comprehensive review for chronic disease prediction using machine learning algorithms" Journal of Electrical Systems and Information Technology: July 2024 DOI:10.1186/s43067-024-00150-4

[4] Mohamed Elhoseny and Rayan Alanazi "Identification and Prediction of Chronic Diseases Using Machine Learning Approach" https://doi.org/10.1155/2022/2826127

[5] Fati Oiza Ochepa1, John Patrick2, Malik Adeiza Rufai3 and Adamu Isah4 "A Deep Learning Based Multiple Chronic Disease Detection Model" NOV 2022 | IRE Journals | Volume 6 Issue 5 | ISSN: 2456-8880

[6] Mohammad Rashedul Islam, Azrin Sultana, Rakibul Islam "A comprehensive review for chronic disease prediction using machine learning algorithms" *Journal of Electrical Systems and Information Technology* volume 11,16 July 2024

[7] Al Khan "Machine Learning for Chronic Disease Prediction". August 05, 2022, CEOS Public. Health. Res. 1(1):101

[8] Dibaba Adeba Debal[1*] and Tilahun Melak Sitote[2] "Chronic kidney disease prediction using machine learning techniques" Debal and Sitote Journal of Big Data (2022) 9:109 https://doi.org/10.1186/s40537-022-00657-5

[9] Sun Min Oh,1,2 Katherine M. Stefani,3 and Hyeon Chang Kim1, "Development and Application of Chronic Disease Risk Prediction Models" Jun 13, 2014. https://doi.org/10.3349/ymj.2014.55.4.853

[10] Kawsher Rahman1*, Prasanna Pasam2, Srinivas Addimulam3 and Vineel Mouli Natakam4 "Leveraging AI for Chronic Disease Management: A New Horizon in Medical Research" Volume 9, No 2/2022 Review Article Malays. j. med. biol. res.

[11] Kosarkar, Gopal Sakarkar, Shilpa Gedam (2022), "An Analytical Perspective on Various Deep Learning Techniques for Deepfake Detection", *1st International Conference on Artificial Intelligence and Big Data Analytics (ICAIBDA),* 10th & 11th June 2022, 2456-3463, Volume 7, PP. 25-30, https://doi.org/10.46335/IJIES.2022.7.8.5

[12] Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2022), "Revealing and Classification of Deepfakes Videos Images using a Customize Convolution Neural Network Model", *International Conference on Machine Learning and Data Engineering (ICMLDE)*, 7th & 8th September 2022, 2636-2652, Volume 218, PP. 2636-2652, https://doi.org/10.1016/j.procs.2023.01.237

[13] Usha Kosarkar, Gopal Sakarkar (2023), "Unmasking Deep Fakes: Advancements, Challenges, and Ethical Considerations", *4th International Conference on Electrical and Electronics Engineering (ICEEE)*,19th & 20th August 2023, 978-981-99-8661-3, Volume 1115, PP. 249-262, https://doi.org/10.1007/978-981-99-8661-3_19

[14] Usha Kosarkar, Gopal Sakarkar, Shilpa Gedam (2021), "Deepfakes, a threat to society", *International Journal of Scientific Research in Science and Technology (IJSRST)*, 13th October 2021, 2395-602X, Volume 9, Issue 6, PP. 1132-1140, https://ijsrst.com/IJSRST219682

[15] Usha Kosarkar, Prachi Sasankar(2021), " A study for Face Recognition using techniques PCA and KNN", Journal of Computer Engineering (IOSR-JCE), 2278-0661,PP 2-5,

[16] Usha Kosarkar, Gopal Sakarkar (2024), "Design an efficient VARMA LSTM GRU model for identification of deep-fake images via dynamic window-based spatio-

temporal analysis", Journal of Multimedia Tools and Applications, 1380-7501, https://doi.org/10.1007/s11042-024-19220-w

[17] Usha Kosarkar, Dipali Bhende, "Employing Artificial Intelligence Techniques in Mental Health Diagnostic Expert System", International Journal of Computer Engineering (IOSR-JCE),2278-0661, PP-40-45, https://www.iosrjournals.org/iosr-jce/papers/conf.15013/Volume%202/9.%2040-45.pdf?id=7557

[18] Usha Kosarkar, Gopal Sakarkar & Mahesh Naik, "A Hybrid Deep Learning Model for robust deep fake detection", *2nd International Conference on Advanced Communications and Machine Intelligence (MICA 2023)*, https://doi.org/10.1007/978-981-97-6222-4_9