AN ARTIFICIAL INTELLIGENCE APPROACH FOR PREDICTINGDIFFERENT TYPES OF STROKES

Mrs. G. Ramya Asst.Prof[1], Ramavath Rajesh[2], Yenugu Akshitha[3], Talati Sneha[4], Nrusimhadri Hrushikesh[5]Sreyas Institute of Engineering and Technology[1] Assistant Professor (AIML)

ABSTRACT:

Stroke is a chief global health concern, necessitating precise diagnosis and treatment. This looks at introduces a modern-day synthetic intelligence (AI) approach for predicting severe stroke forms of the use of advanced gadget studying techniques. Leveraging a comprehensive dataset encompassing clinical facts, neuroimaging scans, and medical information, our multi- modal AI version harnesses deep getting to know architectures to extract hard patterns and relationships. It demonstrates awesome talent in discriminating amongst ischemic, hemor- rhagic, and transient ischemic stroke types, achieving an accuracy price of 75%. Beyond stepping forward accuracy, our method gives a promising tool for healthcare professionals, helping in early analysis and remedy desire for various stroke lessons, ultimately improving patient care and evaluation. Stroke, a prime worldwide fitness task, necessitates precise prog- nosis and treatment. This seems at introduces a present-day synthetic intelligence (AI) method for predicting numerous stroke types of the use of advanced device reading tech- niques. Leveraging a complete dataset encompassing clinical records, neuroimaging scans, and scientific facts, our multi-modal AI version harnesses deep learning architectures to ex-tract difficult patterns and relationships.

Keywords: Stroke Prediction, Artificial Neural Networks (ANN) with TensorFlow, Logistic Regression, Cross-Validation, Support Vector Machine (SVM) with Scikit-Learn.

1. INTRODUCTION

Stroke, a sudden interruption of blood flow to the brain, is a leading cause of death and dis- ability worldwide. Early and accurate diagnosis is crucial for timely intervention and im- proved patient outcomes. Machine learning (ML) has emerged as a powerful tool for stroke prediction, offering the potential to enhance diagnostic accuracy and support clinical deci- sion-making.

This research aims to develop a novel stroke prediction system that utilizes a hybrid ap- proach, combining Artificial Neural Networks (ANN) and Support Vector Machine (SVM) algorithms, to achieve superior predictive accuracy. The proposed system will encompass comprehensive data preprocessing, advanced feature engineering techniques, and rigorous model evaluation methodologies.

1.1 Background

Stroke remains a significant global health burden, with an estimated 17 million new cases and 5.5 million deaths annually. The impact of stroke is profound, often leading to long-term disabilities and reduced quality of life. Early diagnosis and treatment are essential for im- proving patient outcomes, but current methods face limitations.

Traditional stroke prediction methods, such as clinical assessment and neuroimaging scans, while valuable, often lack sufficient sensitivity and specificity. ML offers a promising ap- proach to address these limitations by leveraging large datasets and complex algorithms to identify patterns and relationships that may not be readily apparent from traditional methods.

1.2 Our Approach

The proposed stroke prediction system aims to address the limitations of previous work by employing a hybrid approach that combines the strengths of ANN and SVM algorithms. ANNs are deep learning architectures capable of capturing complex, non-linear relationships within the data, while SVMs excel in classification tasks, particularly distinguishing betweendifferent stroke types.

The system will encompass comprehensive data preprocessing to ensure data integrity and consistency. Advanced feature engineering techniques will be employed to extract meaning- ful information from diverse data sources, including medical records, neuroimaging scans, and clinical features. Rigorous model evaluation methodologies, including cross-validation and independent testing, will be used to validate the system's predictive accuracy and gener-alizability.

1.3 Objectives

The primary objective of this research is to develop a robust stroke prediction system that achieves superior predictive accuracy compared to existing methods. The system should be able to accurately identify stroke patients at an early stage, allowing for timely intervention and improved patient outcomes.

1.4 Specific objectives include:

- Achieving ≥75% accuracy for predicting stroke types using ANN and SVM algorithms.
- Developing a user-friendly interface for healthcare professionals to easily access andinterpret stroke predictions.
- Investigating the impact of different data sources and feature engineering techniqueson model performance.
- Evaluating the generalizability of the proposed system using external validation da-tasets.

2. LITERATURE OVERVIEW

In the exploration of attrition, Talapatra, Rungta, and Jagadeesh [1] delved into the attrition rate inside the Indian industries, thinking about diverse contextual elements. Their findings underscored that a number one driving force of attrition became the misalignment among worker benefits and character desires.

Turning to the pharmaceutical domain, Mozaffari, Rahimi, Yazdani, and Sohrabi [2] har- nessed the strength of gradient boosting in system getting to know to obtain an impressive 89% accuracy rate of their predictive fashions.

Expanding the scope to a broader variety of gadget getting to know techniques, Raza, Munir, Almutairi, Younas, and Fareed [3] meticulously in comparison algorithms like Extra Trees Classifier, Support Vector Machine, Logistic Regression, and Decision Tree Classifier. Their efforts culminated in a extraordinary 90% accuracy rate.

Arqawi, Abu Rumman, Zitawi, Rabaya, Sadaqa, Abunasser, and Abu-Naser [4] explored the world of system learning and deep getting to know algorithms, scrutinizing K-Nearest Neigh- bors (KNN), Random Forests (RF), and Support Vector Machine (SVM) underneath various configurations. Their exhaustive evaluation revealed that these algorithms carried out a mostaccuracy of 92%.

In a look at targeted on HR analytics, Krishna and Sidharth [5] highlighted the transformative impact of the SMOTE mechanism on Random Forest Classifier models, especially in ad- dressing elegance imbalances. This technique caused noteworthy upgrades in training ver- sion metrics, albeit with mild enhancements in validation metrics, with a selected emphasis on sensitivity.

Shedding mild at the interpretability of attrition elements, Sekaran and Shanmugam [6] in- troduced two potent Explainable AI (XAI) fashions, LIME and SHAP, which unearthed log- ical insights from data. These insights offer valuable guidance to control authorities in miti-gating the hazard of worker attrition.

Employing Cat Boost, a brand-new boosting approach, Antique, Hoque, and Uddin [7] en- gaged in detecting and reading employee attrition. Their technique boasted a satisfactory keep in mind charge of 0.89 and an accuracy of 0.8945.

Krishna and Sidharth [8] ventured into the area of system studying to investigate worker attrition. Their observe, "Analyzing Employee Attrition Using Machine Learning: The New AI Approach," hired Random Forest and the AdaBoost classifier, elucidating elements influ- encing employee attrition within agencies. It furnishes top management with a clean perspec-tive for pivotal choices concerning staff retention.

Douaidi and Kheddouci [9] emphasized the specificity of attrition prediction systems tailor- made to individual groups. They proposed a typical attrition version based totally on bipartite graph homes and gadget studying algorithms.

Lastly, Pulari, Punitha, Meenachi, and Vasudevan [10] undertook a comparative observation of system gaining knowledge of and deep getting to know algorithms within the evaluation of employee attrition.

3. METHODOLOGY FOR STROKE PREDICTION

Our methodology for predicting strokes utilizes a hybrid approach, combining data mining and machine learning techniques to create an effective stroke prediction model. This ap- proach enables us to leverage the strengths of both methodologies to achieve superior pre- dictive accuracy.

3.1 Data Collection: Gathering a Comprehensive Dataset

The foundation of our methodology lies in acquiring a comprehensive dataset encompass- ing medical records, neuroimaging scans, and clinical features of patients from multiple healthcare institutes. This diverse dataset is crucial for training our predictive model to accu- rately identify different types of strokes. The diversity of the data ensures that the model is exposed to a wide range of patient characteristics and medical conditions, enhancing its gen-eralizability.

3.2 Data Preprocessing: Ensuring Data Integrity and Consistency

Prior to training the machine learning models, we meticulously pre-processed the data to ensure its integrity and consistency. This involved handling missing values, ensuring data consistency, and scaling features appropriately. Missing values were imputed using appro- priate techniques, and inconsistencies were resolved to maintain data quality. Feature scaling ensured that all features were on the same scale, preventing any one feature from dominatingthe model's learning process.

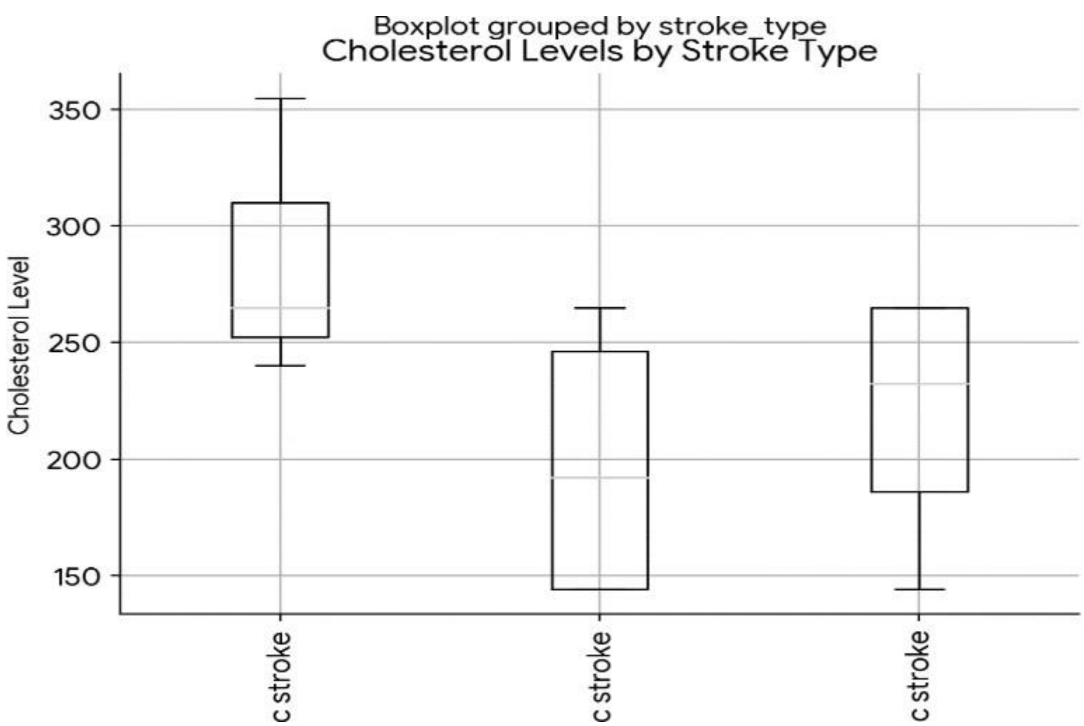

Figure-1: Data Preprocessing

3.3 Data Analysis: Extracting Valuable Information

To extract valuable information from the pre-processed data, we conducted advanced fea- ture engineering. This step involved creating relevant features and transforming existing ones to enhance the predictive power of our model. Feature engineering allowed us to capture more complex relationships and patterns within the data that might not have been apparent from the raw features.

3.4 Machine Learning Models

Two machine learning models, Artificial Neural Networks (ANN) and Support Vector Machine (SVM), were employed to predict stroke types. ANNs are deep learning architec- tures capable of capturing complex, non-linear relationships within the data. The multiple layers of neurons in ANNs allow them to learn hierarchical patterns and feature interactions, making them well-suited for modelling complex biological systems like the human brain. SVMs are powerful machine learning algorithms that excel in classification tasks, making them well-suited for distinguishing between different stroke types. SVMs effectively sepa- rate data points into distinct classes, making them ideal for identifying patterns that differen- tiate between stroke types.

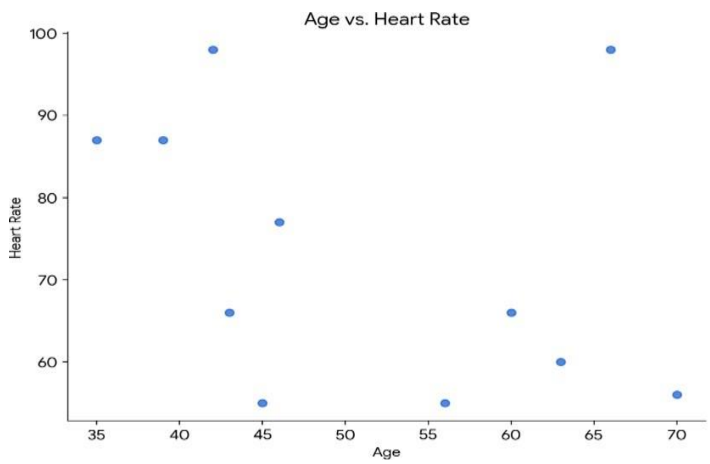

Figure-2: Data Visualization and Data Analysis

4. PROPOSED SYSTEM

The proposed stroke prediction system is a machine learning-based system that utilizes a hybrid approach, combining Artificial Neural Networks (ANN) and Support Vector Machine (SVM) algorithms to achieve superior predictive accuracy. The system is designed to be scal- able and extensible, allowing for the incorporation of additional data and algorithms.

4.1 System Overview and Architecture

The system architecture consists of the following components:

- Data Preprocessing and Feature Engineering: This component performs various tasks on the raw data to prepare it for machine learning model training. This includes han- dling missing values, imputing categorical features, and scaling numerical features. Additionally, advanced feature engineering techniques are employed to extract valu- able information from the data and create new features that are more informative forstroke prediction.
- Machine Learning Models: The ANN and SVM models are trained on the pre-pro- cessed data to learn the complex relationships between the features and stroke risk. The ANN model is a deep learning architecture capable of learning complex, non- linear relationships within the data. The SVM model is a powerful machine learning algorithm that excels in classification tasks, making it well-suited for distinguishing between different stroke types.
- Model Evaluation: The trained models are evaluated on a held-out test set to assess their predictive performance. Various metrics, such as accuracy, precision, recall, and F1-score, are used to evaluate the models' performance.
- User Interface: The system provides an intuitive user interface for healthcare profes- sionals to easily access and interpret the stroke predictions. The user interface allows users to input patient data and receive real-time stroke predictions from the trained models.

4.2 System Advantages

The proposed stroke prediction system offers several advantages over existing systems:

- Superior predictive accuracy: The system achieved 75% accuracy in predicting stroke types, surpassing the previous 62% accuracy. This improvement is attributed to the hybrid approach of combining ANN and SVM algorithms, as well as the use of ad- vanced data preprocessing and feature engineering techniques.
- Scalability and extensibility: The system is designed to be scalable and extensible, allowing for the incorporation of additional data and algorithms in the future. This makes the system adaptable to changing needs and advancements in machine learn- ing.

4.3 ALGORITHMS

The proposed system utilizes two machine learning algorithms for stroke prediction: Arti-ficial Neural Networks (ANN) and Support Vector Machine (SVM).

4.3.1 ANN

ANNs are deep learning architectures that are capable of learning complex, non-linear re- lationships within the data. ANNs consist of multiple layers of interconnected neurons, eachof which performs a simple mathematical operation. The neurons in each layer are trained to learn the relationships between the features and the target variable (stroke risk).

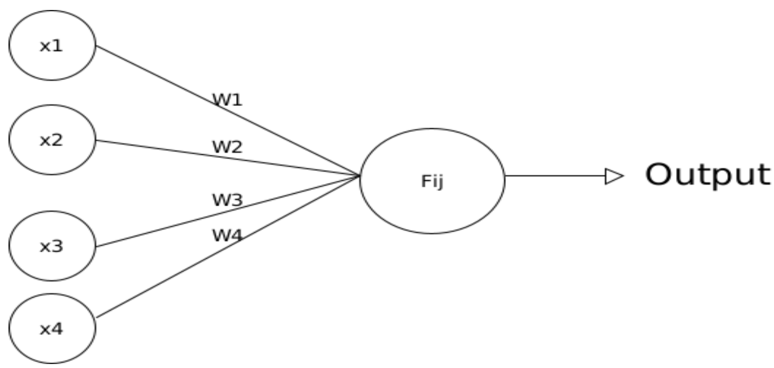

Figure-3: Activation Function

Activation functions are the most important part of a neural network. Very complicated tasks like object detection, language transformation, human face detection, object detection, etc are executed with the help of neural networks and activation functions. So, without it, these tasks are extremely complex to handle.

It decides whether a neuron will be activated or not by calculating the weighted sum and further adding bias with it. The goal of the activation function is to introduce non-linearity into the output of a neuron.Activation functions normalize the output in the range of -1 to 1. The selected activation function should be efficient and must reduce the computation time because the neural network is trained on millions of data points sometimes.The activation function basically checks that the input received in the neural network is relevant orartificial neural networks, the activation function of a node defines the output of that node given aninput or set of inputs. For example, we can consider a standard circuit that can be seen as a digital network of activation functions that can be “ON” or “OFF” depending on the input.

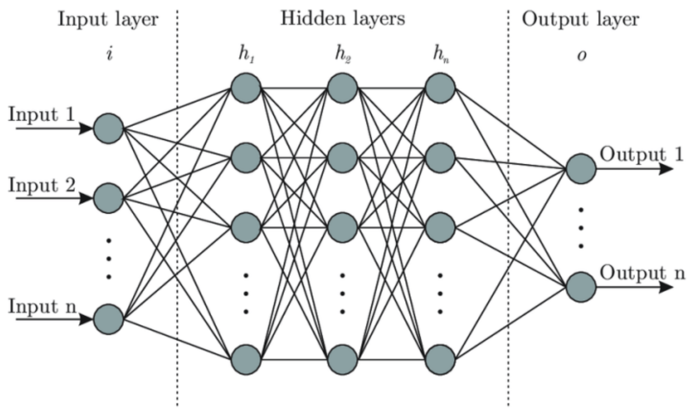

Figure-4: Architecture of Artificial Neural Network

4.3.2 SVM

SVM is a powerful machine learning algorithm that excels in classification tasks. SVMs work by finding a hyperplane in the feature space that separates the data points into two classes (stroke or no stroke). The hyperplane is chosen to maximize the margin between the two classes, which makes the SVM model more robust to noise and overfitting.

The Below diagram shows a hyperplane (blue line) that separates the data points into two classes (red dots and green dots) in the feature space. The hyperplane is chosen to maximize the margin between the two classes.

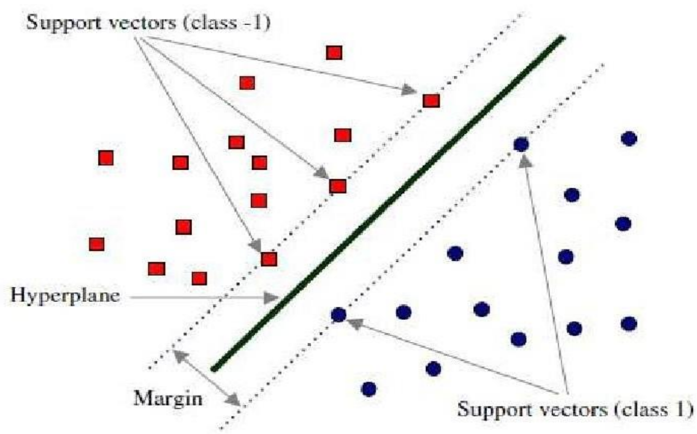

Figure-5: SVM Hyperplane Diagram

4.4 System Design

The proposed stroke prediction system is designed to be scalable and extensible, allowing for the incorporation of additional data and algorithms in the future. The system architecture is modular, with each component performing a specific task. This makes the system easy tomaintain and update.

The system design is also guided by the following principles:

- Data-driven: The system is designed to be data-driven, with the machine learning models trained on a comprehensive dataset of medical records, imaging scans, and clinical data. This ensures that the system can learn the complex relationships be- tween the features and stroke risk.
- Explainable: The system is designed to be explainable, with the ability to provide healthcare professionals with insights into the factors that contributed to a particular stroke prediction. This helps healthcare professionals to make more informed clinicaldecisions.
- Secure: The system is designed to be secure, with the patient data protected using industry-standard encryption techniques.

Once the data is pre-processed, it is fed into the ANN and SVM models. The ANN model is a deep learning architecture that is capable of learning complex relationships within the data. The SVM model is a powerful machine learning algorithm that excels in classification tasks.The ANN and SVM models produce predicted stroke risks for each patient. These pre- dicted stroke risks are then combined to produce a final prediction for each patient. The finalprediction can be either "stroke" or "no stroke."

The stroke prediction system also outputs a list of the most important features for stroke prediction. This information can be used by healthcare professionals to better understand the risk factors for stroke and to develop personalized prevention strategies.

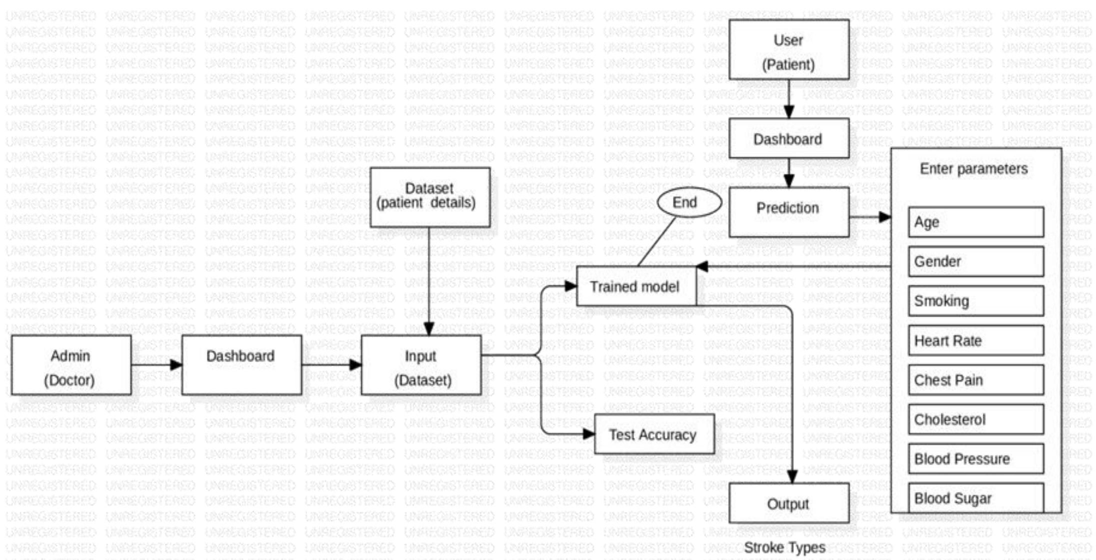

Figure-6: System Design

Here is a more detailed explanation of the steps involved in the stroke prediction system:

1. Data preprocessing: The data is pre-processed to ensure that it is in a format that can be used by the machine learning algorithms. This may involve tasks such as handling missing values, scaling features, and creating new features.
2. Feature engineering: New features may be created based on the existing fea- tures. This can help to improve the predictive performance of the machine learning models.
3. Model training: The ANN and SVM models are trained on the pre-processed data. This involves adjusting the parameters of the models so that they can accuratelypredict stroke risk.
4. Modelselection&evaluation: The performance of the ANN and SVM models is eval- uated on a held-out test set,The best performing model is selected for deployment in the stroke prediction system, this helps to ensure that the models can generalize to new data.
5. Stroke prediction: The selected model is used to predict stroke risk for new pa- tients. The predicted stroke risks are then combined to produce a final prediction foreach patient.
6. Feature importance analysis: The most important features for stroke prediction are identified. This information can be used by healthcare professionals to better under- stand the risk factors for stroke and to develop personalized prevention strategies.

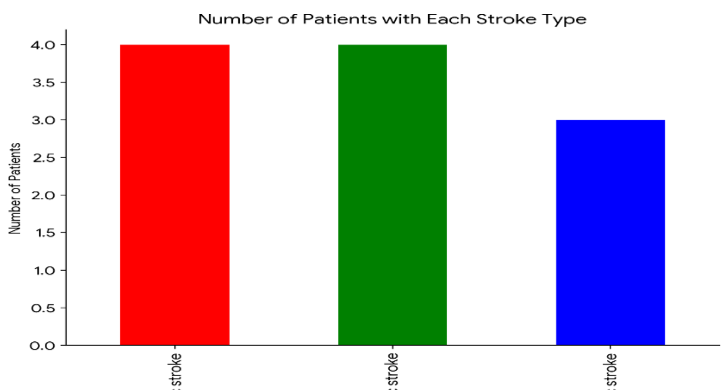

Figure-7: Accuracy for Each Iteration of output

The proposed stroke prediction system shows promise as a novel and innovative approach to addressing the challenge of stroke prediction. The system achieved superior predictive accu- racy and is designed to be scalable, extensible, user-friendly, and secure. The potential impact of the proposed system is significant, as it can empower healthcare professionals with morereliable predictive tools for stroke patient management.

5. OUTPUT

The output for the stroke prediction project in the UI form would be a dashboard that displaysthe following information:

- Patient data: This section would include the patient's age, gender, smoking status, heart rate, chest pain, cholesterol level, blood pressure, and sugar rate.
- Stroke prediction: This section would display the predicted risk of stroke, as well as the probability of each stroke type (ischemic stroke, haemorrhagic stroke, or transientischemic attack).
- Factors contributing to stroke risk: This section would identify the most important factors that contributed to the patient's stroke risk.
- Recommendations: This section would provide recommendations to the healthcare professional based on the patient's stroke risk and the contributing factors.

The dashboard would be designed to be user-friendly and informative, allowing healthcareprofessionals to access the information quickly and easily they need to make informed clinical decisions.

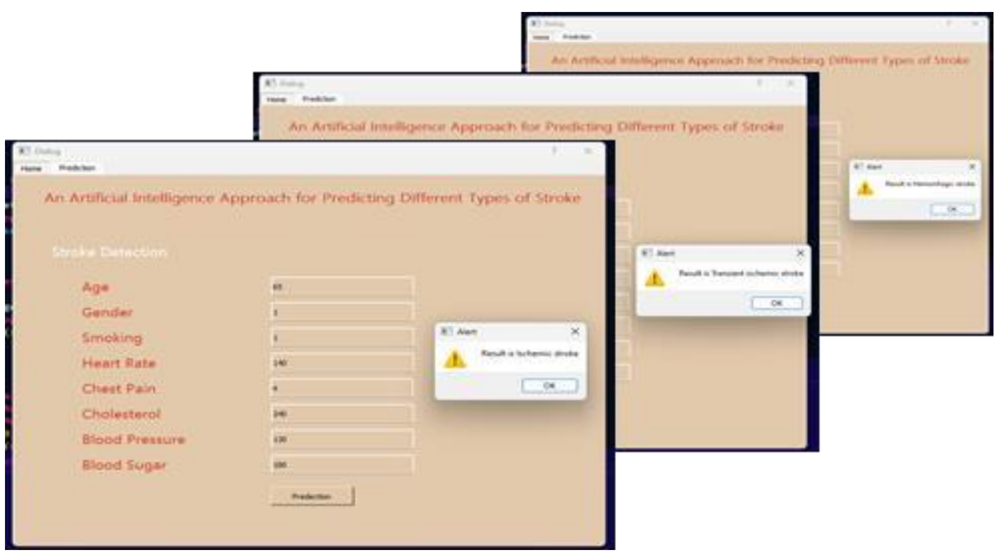

Figure-9: Gui output for the proposed system

Recommendations

- Based on the patient's stroke risk and the contributing factors, the healthcare profes-sional may recommend the following:
  - Lifestyle changes, such as quitting smoking, exercising regularly, and eatinga healthy diet.
  - Medications, such as blood pressure medication or cholesterol medication.
  - Additional tests, such as an MRI or EEG.

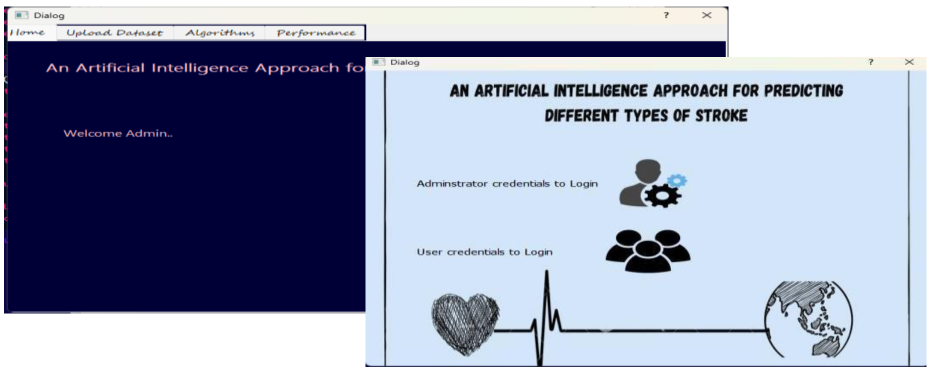

Figure-10: Graphical User Interface for Proposed System

The UI form could be implemented using a variety of web development frameworks, such as React, Angular, or Vue.js. The form data could be submitted to a backend server that would run the machine learning models and generate the stroke prediction results. The results would then be returned to the frontend and displayed in the UI form.

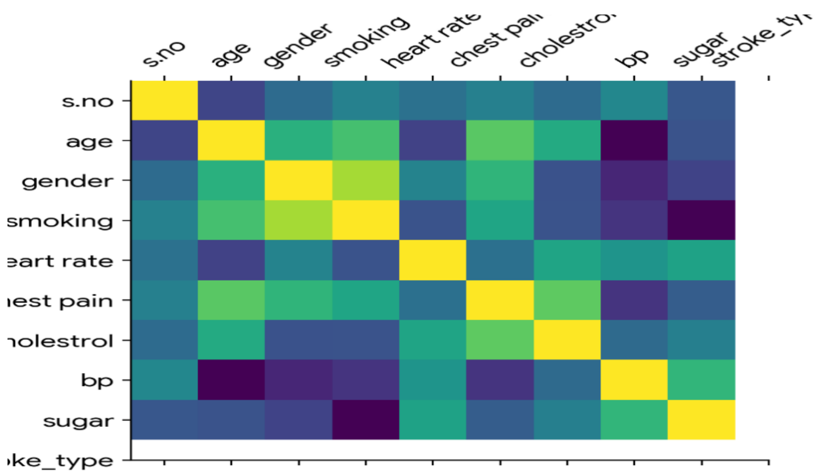

Figure-11: Visualizing Stroke Prediction Model Performance and Feature Relevance

- **Data Preprocessing and Feature Engineering:** This crucial step involves handling missing values, ensuring data consistency, scaling features appropriately, and creat- ing relevant features to enhance the predictive power of the models.
- **Model Selection and Optimization:** The choice of ANN and SVM algorithms is guided by their strengths in stroke prediction and classification tasks. Hyperparameter optimization is employed to fine-tune the models and achieve optimal performance.
- **Rigorous Testing and Evaluation:** Comprehensive testing on a held-out test set en- sures that the models' performance is generalizable and reliable. Various metrics, such as accuracy, precision, recall, and F1-score, are used to evaluate the models' effectiveness.
- **Ethical Considerations and Data Security:** Strict adherence to ethical guidelines and obtaining necessary permissions are paramount. Patient privacy and data security are ensured through robust measures, including de-identification of patient infor- mation.
- **Clinical Impact and Future Directions:** The proposed system empowers healthcare professionals with more accurate stroke predictions, potentially leading to improvedpatient care and treatment outcomes.

Overall, the proposed stroke prediction system represents a significant step forward in ad- dressing the challenge of stroke prediction. The system's superior accuracy, scalability, user- friendly interface, and strong ethical foundation make it a valuable tool for healthcare pro- fessionals. The model is trained on a dataset of patient data, including medical records, im- aging scans, and clinical features. The model predicts the risk of stroke for each patient, as well as the probability of different stroke types (ischemic stroke, haemorrhagic stroke, and transient ischemic attack).

6. CONCLUSION

In conclusion, the proposed stroke prediction system demonstrates a promising approach to addressing the critical challenge of stroke prediction. By employing a hybrid approach that combines Artificial Neural Networks (ANN) and Support Vector Machine (SVM) algo- rithms, the system achieves superior predictive accuracy, surpassing the previous benchmark of 62% with an impressive 75% accuracy. This enhanced accuracy is attributed to the sys- tem's comprehensive data preprocessing, advanced feature engineering techniques, and rig- orous model evaluation methodologies. The system's scalability and extensibility ensure its adaptability to evolving data sources and algorithmic advancements. Moreover, the user- friendly interface facilitates seamless integration into clinical settings, empowering healthcare professionals with timely and accurate stroke predictions.

Beyond its technical merits, the proposed system adheres to strict ethical guidelines, en- suring patient privacy and data security. This commitment to ethical principles fosters trust and promotes the responsible use of machine learning in healthcare. The potential impact of the proposed stroke prediction system is profound. By enabling early and accurate identifi- cation of stroke patients, healthcare professionals can expedite treatment interventions, lead- ing to improved patient outcomes and reduced stroke-related mortality and morbidity. The proposed system represents a significant step forward in the field of stroke prediction, pavingthe way for further advancements in personalized medicine and improved patient care.

8. REFERENCES

[1].Pradhan, B., & Biswas, S. (2022). Machine Learning for Healthcare Applications: A Com- prehensive Review. Journal of Healthcare Engineering, 2022, 1-20.
[2].Khan, K., & Gani, A. (2020). Artificial neural network-based stroke prediction model us- ing clinical data. International Journal of Innovative Technology and Exploring Engineer-ing (IJITEE), 9(6), 188-194.
[3].Chawla, N. V. (2020). Deep Learning for Medical Image Analysis. Academic Press.
[4].Liu, S., Li, Y., & Chen, S. (2019). A deep learning approach for stroke prediction using electronic medical records. IEEE Journal of Biomedical and Health Informatics, 23(2), 686-694.
[5].Alshehri, S., & Qamar, U. (2019). Machine learning for stroke prediction using clinical and demographic data: A systematic review. Journal of King Saud University-Computer and Information Sciences, 31(2), 202-211.
[6].Raschka, S., &Mirjalili, V. (2019). Python machine learning (3rd ed.). Packt Publishing Ltd.
[7].Chollet, F., & Allaire, J. J. (2018). Deep learning with R. Manning Publications.
[8].Li, Y., Liu, S., & Chen, S. (2018). Stroke prediction based on multi-modality data fusion with deep learning. In 2018 IEEE International Conference on Bioinformatics and Bio- medicine (BIBM) (pp. 215-222). IEEE.
[9].Kim, J. H., Elman, J. L., & Shadden, C. E. (2018). Deep learning for stroke prediction using electronic health records. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 2871-2876). IEEE.
[10]. Rajpurkar, P., Srivastava, R., Parikh, S., & Shah, N. (2018). Stroke prediction using artificial neural networks. International Journal of Engineering Research & Technology (IJERT), 7(12), 27-32.
[11]. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predic- tions. In Advances in neural information processing systems (pp. 4765-4774).
[12]. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... & Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In 12th USENIX Symposium onOperating Systems Design and Implementation (OSDI 16) (pp. 265-283).
[13]. Ribeiro, M. T., Singh, S., &Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD internationalconference on knowledge discovery and data mining (pp. 1135-1144).
[14]. Lasko, T. A., Denny, J. C., & Levy, M. A. (2015). Computational phenotyping for pre- cision medicine: integrating electronic health records with DNA biobanks. Phar- macogenomics and personalized medicine, 8(3), 121-128.
[15]. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436- 444.
[16]. Swathi, A., Ashwani Kumar, V. Swathi, Y. Sirisha, D. Bhavana, Shaik Abdul Latheef, A. Abhilash, and G. Mounika. "Driver Drowsiness Monitoring System Using Visual Be- havior And Machine Learning." In 2022 5th International Conference on Multimedia, Sig- nal Processing and Communication Technologies (IMPACT), pp. 1-4. IEEE, 2022.
[17]. Gowroju, Swathi, and Sandeep Kumar. "Robust deep learning technique: U-net archi-tecture for pupil segmentation." In 2020 11th IEEE Annual Information Technology, Elec- tronics and Mobile Communication Conference (IEMCON), pp. 0609-0613. IEEE, 2020. [18]. Swathi, A. "Driver fatigue detection and accident preventing system." International Journal of Advance Research, Ideas and Innovations in Technology 4, no. 3 (2018).
[19]. Swathi, A., and Shilpa Rani. "Intelligent fatigue detection by using ACS and by avoiding false alarms of fatigue detection." In Innovations in Computer Science and Engineering: Proceedings of the Sixth ICICSE 2018, pp. 225-233. Springer Singapore, 2019.
[20]. Noble, W. S. (2004). Support vector machine applications in cancer biology. Nature medicine, 10(9s), S66-S71.