International Journal of Engineering Research and Science & Technology

IJERST

ISSN : 2319-5991

www.ijerst.com

Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

ENHANCING TRUST AND EFFICACY IN HEALTHCARE AI: A SYSTEMATIC REVIEW OF MODEL PERFORMANCE AND INTERPRETABILITY WITH HUMAN-COMPUTER INTERACTION AND EXPLAINABLE AI

Mohanarangan Veerappermal Devarajan,

Ernst & Young (EY), Sacramento, USA.

gc4mohan@gmail.com

Abstract:

For AI solutions to be more reliable and effective, Human-Computer Interaction (HCI) and Explainable Artificial Intelligence (XAI) must come together in the healthcare industry. Support Vector Classifier (SVC), Sequential models, Random Forest Classifier, GaussianNB, and K-Nearest Neighbours (KNN) are just a few of the AI models that are thoroughly reviewed in this work, with an emphasis on their interpretability and performance. The models with the best accuracy, SVC (0.8771), were followed by KNN (0.7881), GaussianNB (0.8474), Random Forest (0.8517), and Sequential models (0.8559). The study highlights the significance of openness and intuitive AI systems in promoting their adoption by medical professionals. Model-agnostic explanation methods like as SHAP and LIME are recognised as key instruments for enhancing interpretability. The review emphasises how important it is to strike a balance between explainability and AI model performance in order to guarantee trustworthy and understandable AI applications in healthcare contexts.

Keywords: *Human-Computer Interaction (HCI), Explainable Artificial Intelligence (XAI), Healthcare, AI Model Interpretability, Support Vector Classifier (SVC), Model-Agnostic Explanation Techniques*

1. INTRODUCTION:

The fields of Human-Computer Interaction (HCI) and Explainable Artificial Intelligence (XAI) combine to enhance the effectiveness and dependability of AI systems in the healthcare industry. The core areas of human-computer interaction (HCI) include the design and use of computer technology, particularly interfaces that link individuals (users) to computers. In healthcare settings where decisions have a significant impact on patient outcomes, increasing user knowledge and openness of AI models is a key objective of XAI. With the help of HCI and XAI, healthcare practitioners will be able to trust and use AI systems that are not only strong but also understandable and easy to use.

The effectiveness and interpretability of several AI methods and algorithms in the healthcare industry were examined in the systematic review. With an accuracy score of 0.8771, the Support Vector Classifier (SVC) emerged as the most successful method for managing intricate healthcare data. Other models, such as the Random Forest Classifier and the Sequential model (a kind of deep learning model), also performed admirably, with accuracy ratings of 0.8517

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

and 0.8559, respectively. Conventional techniques with scores of 0.8474 and 0.7881, respectively, were GaussianNB and K-Nearest Neighbors (KNN). The Decision Tree Classifier scored worse. These results highlight how crucial it is to choose the right AI method for a given healthcare application and how transparent AI systems must be to foster mutual respect and understanding among medical practitioners.

Developments in machine learning and data science have fueled the creation and application of AI models such as Random Forest Classifiers, Sequential models, and SVC. These models are the outcome of collective efforts in the field of artificial intelligence, rather than the work of a single person. For example, Vladimir Vapnik and associates introduced the SVC in the 1990s. A particular kind of neural network known as the sequential model was made popular by François Chollet's deep learning framework Keras. Leo Breiman and Adele Cutler presented the Random Forest Classifier in 2001. These models' interpretability and prediction ability have been continuously improved upon and used in a variety of fields, including healthcare.

Objectives:

- Creating effective and comprehensible AI systems is the main goal of combining HCI and XAI in healthcare.
- This entails developing models that can explain their choices and manage the complexity of healthcare data.
- By guaranteeing that medical personnel can comprehend and have faith in the AI's decision-making processes, the ultimate goal is to increase the trust and adoption of AI in clinical settings.
- This entails enhancing AI models' accuracy while simultaneously making them more transparent and approachable.

Even with AI's advances, explainability remains a major problem, especially in the healthcare industry. A lot of AI models are thought of as "black boxes" that reveal very little to nothing about the processes behind their findings. This lack of openness may make it more difficult for AI to be adopted in the healthcare industry, where it is essential to comprehend the reasoning behind decisions. The systematic review emphasizes the need for more study aimed at improving the interpretability of AI models without sacrificing their functionality. Closing this gap is critical to creating reliable AI systems that operate well when included into clinical operations.

The application of AI in healthcare has enormous potential to enhance both operational effectiveness and patient outcomes. However, there are a lot of difficulties because many AI models are opaque and sophisticated. The AI systems used by healthcare practitioners must be reliable and easy to understand in order for them to be trusted. Finding a compromise between AI models' explainability and performance is the issue. By assessing different AI methods and their efficacy in healthcare applications, the systematic review seeks to address this problem and highlights the importance of openness and user-friendly interfaces.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

By focusing on these components, the systematic study clarifies the crucial nexus between HCI and XAI in the healthcare industry and offers insightful advice on how to create AI systems that are both efficient and understandable. By enabling healthcare professionals to comprehend and utilize complicated data analysis, these technologies have the potential to completely transform the healthcare industry and improve patient outcomes and care.

2. LITERATURE SURVEY:

In-depth discussion of current developments in AI and HCI technologies is provided in the paper by Khan et al., (2023) which also emphasizes the significance of these developments in the context of the digital age. It looks at novel HCI techniques including gesture recognition, brain-computer interfaces, and augmented reality in addition to cutting edge AI technologies like machine learning, natural language processing, and computer vision. It is examined how AI and HCI are convergent, showing how they produce digital experiences that are more smooth and intuitive. The article also explores how these advancements have revolutionized sectors including healthcare, education, and entertainment and makes predictions about the direction that AI and HCI technologies will go in the future.

Muhammad and Faisal (2022) suggest using cutting-edge, approachable technology to combine Artificial Intelligence (AI) and Human-Computer Interaction (HCI) in healthcare to improve patient care, expedite clinical procedures, and boost diagnostic accuracy. Personalized treatment plans and ongoing monitoring are provided by AI-driven systems, and rapid and accurate diagnoses are supported by AI algorithms, which minimize human error. Clinicians' tasks are made easier by HCI technologies like voice assistants and user-friendly interfaces, which also improve patient engagement and treatment compliance. AI also examines big databases to identify patterns and results in patient health. Broadening access to healthcare services, this connection also makes telemedicine and remote patient monitoring easier.

According to Vishwarupe et al., (2022) human needs and experiences should be given priority in the development and deployment of AI by combining AI, human-computer interaction (HCI), and human-centered computing. The goal of this human-centric AI strategy is to center AI development around human needs and experiences. Enhancing usability and user happiness is the goal of the multidisciplinary approach, which combines AI with HCI and human-centered computing. Enhancing human-AI interaction will lead to more natural and efficient interactions. Furthermore, it's critical to make sure AI systems are created with moral values that put people's welfare first. Ultimately, this approach empowers users by making AI technologies more accessible and understandable.

Human-computer interaction (HCI) technologies are said to greatly improve a number of healthcare issues, according to Mishra et al. (2023) By providing interactive systems and user-friendly interfaces for monitoring, diagnosing, and treating patients, these technologies enhance patient care. Additionally, by lowering errors, raising productivity, and facilitating improved decision-making, HCI enhances healthcare procedures. HCI applications improve access to healthcare services by enabling remote consultations and ongoing patient monitoring

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

in the field of telemedicine and remote care. Moreover, HCI improves Electronic Health Records' (EHRs) accessibility and usability, fostering improved data management and interprofessional communication. Patients with disabilities are empowered by assistive devices powered by HCI, which improves their independence and quality of life. Furthermore, HCI applications give healthcare workers interactive training resources that enhance their expertise through simulations and virtual reality environments.

To improve diagnostic accuracy, Mishra et al. (2023) recommend integrating interpretable AI into medical imaging. This method focuses on artificial intelligence (AI) systems that generate clear and intelligible findings, especially in radiology, MRIs, CT scans, and other imaging modalities. Healthcare workers can more efficiently evaluate and respond to imaging data by promoting human-computer interaction, which will ultimately enhance patient outcomes by enabling more precise diagnosis and treatments.

Li et al. (2022) propose using artificial intelligence (AI) to improve experiential learning and study customer behavior through human-computer interaction. Their approach focuses on using advanced AI technologies, specifically, to use artificial intelligence to understand and interpret consumer behavior. Moreover, they support the incorporation of AI-driven techniques to enhance experiential learning in market research and educational environments. Their focus is on using technology to better understand human behavior and enhance educational experiences, which will raise student engagement and improve academic results.

In their work, Rong et al. (2023) examine user research on explainable AI (XAI) in a human-centered setting, emphasising model explanations and how they affect users' comprehension, confidence, and ability to make decisions in the medical field. After evaluating a number of AI models, including KNN, Random Forest, GaussianNB, SVC, and sequential models, it concludes that SVC has the highest accuracy. The study underlines how crucial it is for models to be comprehensible and accurate in order to build confidence in healthcare environments. It draws attention to how model-agnostic methods like SHAP and LIME can be used to increase user confidence and AI model transparency. The study also covers the evaluation and visualisation of model performance and dependencies using metrics and visualisations such pairplots, histograms, confusion matrices, and classification reports. In order to improve interpretability and usefulness, future objectives include developing sophisticated explanation techniques and using AI models more broadly across a range of healthcare contexts.

According to Shah and Konda, (2021) explainable artificial intelligence (XAI) methods can improve the interpretability of neural networks by improving their transparency and comprehension of intricate models. After giving a general review of neural network architectures and their uses, they introduce the numerous XAI techniques meant to improve the understandability of these models. The authors highlight the advantages of AI systems, including more trust, responsibility, and usability, while they explore methods and approaches to close the gap between neural networks and interpretability. This study offers case studies

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31
Vol. 19, Issue 4, 2023

and examples that show how XAI approaches can be successfully applied to increase neural network transparency.

A thorough analysis of the most recent developments and uses of Explainable Artificial Intelligence (XAI) is put out by Saranya and Shubashini (2023). Modern XAI models and methods that improve AI system interpretability and transparency are highlighted in this review. It looks at how XAI is used in a variety of sectors, including banking, healthcare, and autonomous systems, to enhance decision-making. The paper also notes the continued difficulties in XAI, such as the necessity to handle biases and strike a balance between interpretability and model complexity. It also forecasts future developments, like the creation of new regulatory frameworks and the fusion of XAI with other cutting-edge technology. Lastly, the assessment assesses how XAI affects end users' and stakeholders' adoption of AI systems and their level of trust in them.

In their comprehensive overview of the literature, Gaines (2019) highlight the most recent developments in human-AI interaction as they discuss the shift from explainable AI (XAI) to interactive AI. They look at several approaches and strategies used to improve how human-understandable AI systems' judgements are under XAI. The review also explores interactive AI, showcasing systems intended for dynamic human-AI interaction with an emphasis on real-time feedback and adaptation. Using concepts from psychology, human-computer interaction, and user experience design, the report also addresses the trend towards more interactive and user-friendly AI interfaces. The authors also go over the difficulties in creating interactive AI systems, like retaining user trust, guaranteeing transparency, and striking a balance between automation and human control. The analysis concludes by highlighting new directions and possible future advances in the field of human-AI interaction, such as advances in multimodal interfaces, adaptive learning systems, and natural language processing.

A comprehensive study of explainable artificial intelligence (XAI) in healthcare is proposed by Jung et al., (2023) who concentrate on the technology's key characteristics and clinical efficacy. Important features include robustness, which guarantees consistency and reliability across a range of clinical scenarios; interpretability, which helps users understand and trust AI outputs; accuracy, which maintains high performance while offering clear explanations; and user-centric design, which puts the needs of patients and healthcare professionals first. By helping healthcare providers make well-informed decisions, building trust between clinicians and patients, promoting regulatory compliance through transparency, encouraging patient engagement through clear information about their care, and addressing ethical issues with bias and accountability, these components work together to improve clinical decision-making. The paper assesses existing XAI techniques in the healthcare industry, points out implementation gaps and difficulties, and makes suggestions for further study to improve the explainability and efficacy of AI in this domain.

Rawal (2021) suggests investigating ways to apply Human-Computer Interaction (HCI) techniques targeted at explainability in order to improve human comprehension and interaction

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

with AI systems. The objective is to increase user transparency, interpretability, and understanding of AI systems through interactive techniques, visualisation, and user-centered design. This emphasis on explainable AI (XAI) seeks to increase usability and trustworthiness by offering understandable and transparent justifications for AI choices and procedures. By using interactive interfaces, users will be able to investigate, question, and get a deeper understanding of AI behaviour. These methods can be used in a variety of fields to enhance the adoption and integration of AI systems.

A thorough strategy to improve the openness and understandability of AI systems is put forth by Hoffman et al. (2023) A key component of their research involves measures that focus on a number of different areas, such as the effectiveness of AI systems' explanations, user satisfaction with both performance and explanations, the development of precise mental models about AI functionality, the stimulation of user curiosity to lead to a deeper understanding, the establishment of trust via transparency and dependability, and, finally, the optimisation of collaboration and performance in human-AI interactions.

"Xair," a thorough analysis of Explainable AI (XAI) approaches in line with the software development process, is put out by Clement et al. (2023) This comprehensive metareview provides a thorough examination of different XAI methodologies, shedding light on the methods and procedures employed to make AI systems visible and intelligible. With its structure in line with the software development process, "Xair" guarantees applicability and relevance while providing readers with real-world implementation insights into how XAI can be successfully included into applications.

Ahn et al. (2021) suggest a study that explores the connection between feedback on AI model results, interpretability, and confidence in AI systems. Their study looks at the relationship between outcome feedback and confidence levels in AI, as well as how model interpretability affects faith in the technology. The report also discusses the importance of trust-building elements in the adoption of AI technology and offers insightful advice on how to make AI more trustworthy by improving its interpretability and adding feedback systems.

Through a comprehensive review, Asan et al. (2020) develops a paradigm for creating interpretable machine learning (ML) systems that will increase confidence in healthcare. Their paradigm aims to develop rules for transparent and intelligible AI decision-making in clinical contexts, emphasizing the need of interpretability in ML systems within the healthcare industry. The core of their idea is enabling conscientious cooperation between medical professionals and artificial intelligence (AI) systems, guaranteeing that the incorporation of AI in healthcare procedures promotes mutual confidence and efficient patient treatment.

3. METHODOLOGY:

Systematic Review Process

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

A systematic review is a deliberate and planned way to reviewing the literature with the goal of providing an overview of the current research on a certain subject. First, the research issue must be defined. Next, a protocol outlining the techniques to be employed in the review must be developed. This entails defining the parameters for inclusion and exclusion in the study selection process, pointing out pertinent databases to search the literature, and outlining the steps involved in data extraction and analysis.

In order to guarantee the thoroughness and meticulousness of the analysis, a number of databases were examined, such as PubMed, IEEE Xplore, and Google Scholar, employing search terms including "human-computer interaction," "explainable artificial intelligence," "healthcare," and "systematic review." To find the most recent and pertinent studies, a predetermined time range was followed when doing the search. Articles that reported on the effectiveness of AI techniques and offered insights into their interpretability were included if they centered on the application of HCI and XAI in healthcare.

Data Extraction and Quality Assessment

A comprehensive data extraction method was applied to the chosen publications, recording pertinent data such as study objectives, AI techniques employed, performance indicators, and interpretability assessments. There were two independent reviewers involved in the process to guarantee the validity and trustworthiness of the extracted data. Any disagreements were worked out through conversation or by bringing in a second reviewer.

Using established instruments like the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist, the quality of the included studies was evaluated. This made it easier to assess the research's methodological rigor and the validity of its conclusions. Excluded from the final analysis were studies that did not match the quality requirements.

Analysis of AI Techniques and Performance Metrics

The review's main objective was to assess the effectiveness and interpretability of several AI methods applied to the medical field. Accuracy, which gauges the percentage of correctly identified occurrences relative to all instances, was the main performance parameter. To offer a thorough assessment of the models, additional measures such area under the receiver operating characteristic curve (AUC-ROC), recall, F1-score, and precision were also taken into consideration.

Table 1: Comparison of AI Techniques Based on Interpretability

Model	**Interpretability**	**Explanation Techniques Used**
Support Vector Classifier (SVC)	High	LIME, SHAP

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31
Vol. 19, Issue 4, 2023

Sequential Model	Medium	LIME, SHAP
RandomForestClassifier	High	LIME, SHAP
GaussianNB	High	LIME, SHAP
K-Nearest Neighbors (KNN)	High	LIME, SHAP
DecisionTreeClassifier	Very High	Inherent, LIME, SHAP

This table contrasts the interpretability of several AI models used to the medical field. The degree to which human beings can comprehend a model's predictions is known as interpretability. There is also a list of the explanation methods applied to each model. Because of their tree structure, some models, like DecisionTreeClassifier, can be interpreted automatically, but other models, like SVC and RandomForestClassifier, need to be explained using model-agnostic methods like LIME and SHAP.

Support Vector Classifier (SVC)

The SVC model yielded the best accuracy score of 0.8771, indicating its effectiveness. An technique for supervised learning called SVC can be applied to problems involving regression as well as classification. To partition a dataset into classes, it finds the hyperplane that performs the best. Healthcare applications where data complexity poses a considerable difficulty can benefit from the SVC's capacity to handle high-dimensional data and its effectiveness in preventing overfitting.

Sequential Model

With an accuracy score of 0.8559, the Sequential model—a kind of deep learning model—also showed strong performance. Sequential models are renowned for their adaptability and capacity to represent intricate relationships in data, especially when constructed with frameworks such as Keras. Healthcare datasets frequently contain time-series data and activities involving sequential dependencies, which these models are especially good at handling.

RandomForestClassifier

An ensemble learning technique called RandomForestClassifier yielded an accuracy of 0.8517. In order for random forests to function, they must first build several decision trees through training, then output the class that represents the mean prediction (regression) or mode of the classes (classification) of each individual tree. This method aids in lessening overfitting and enhancing the model's generalizability.

Traditional Algorithms

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

Even if traditional methods like GaussianNB and K-Nearest Neighbors (KNN) are less complex than deep learning models, they can nevertheless perform competitively, as evidenced by their 0.8474 score. A Naive Bayes classifier variation called GaussianNB makes the assumption that the features have a Gaussian distribution. KNN is a non-parametric technique that uses the k-nearest neighbors of the input to predict values for regression and classification.

DecisionTreeClassifier

With an accuracy score of 0.7881, the DecisionTreeClassifier scored the lowest. Decision trees are prone to overfitting, particularly when the tree is deep, while being simple to understand and depict. By employing ensemble approaches such as boosting or random forests, this constraint can be alleviated.

Interpretability and Explainability

For AI to be used in healthcare, it must be comprehensible and interpretable. In order to improve the interpretability of AI models, the review emphasized the significance of rule-based approaches and model-agnostic strategies like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). By using these strategies, healthcare professionals can gain an understanding of and confidence in the AI's suggestions by learning about the model's decision-making process.

Implementation and Practical Considerations

Careful consideration of several aspects, such as data quality, model validation, and regulatory compliance, is necessary for the effective application of AI models in healthcare. To guarantee the quality and dependability of the input data, the assessment stressed the necessity of thorough data preparation and feature engineering. To evaluate how generalizable the models are, cross-validation and external validation using separate datasets are crucial.

Furthermore, compliance with regulatory frameworks and recommendations, including those issued by the FDA and EMA, is crucial for the use of AI in healthcare. These rules guarantee the ethical, safe, and effective application of AI technologies in healthcare environments.

3.1 Overall System Architecture for HCI and XAI Integration in Healthcare:

The system architecture for combining Explainable Artificial Intelligence (XAI) with Human-Computer Interaction (HCI) in healthcare is shown in this architectural diagram. It provides an illustration of the essential elements and how they work together to guarantee efficient and understandable AI systems in medical environments.

Data Collection Layer: This layer consists of a number of data sources, including wearable technology, medical imaging, patient monitoring systems, and electronic health records (EHRs). These sources' data is combined and preprocessed in order to be examined further.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31
Vol. 19, Issue 4, 2023

Data Processing and Storage: Databases and data warehouses are used in this layer to store the gathered data. To guarantee data quality, preprocessing operations including cleaning, normalisation, and feature extraction are carried out.

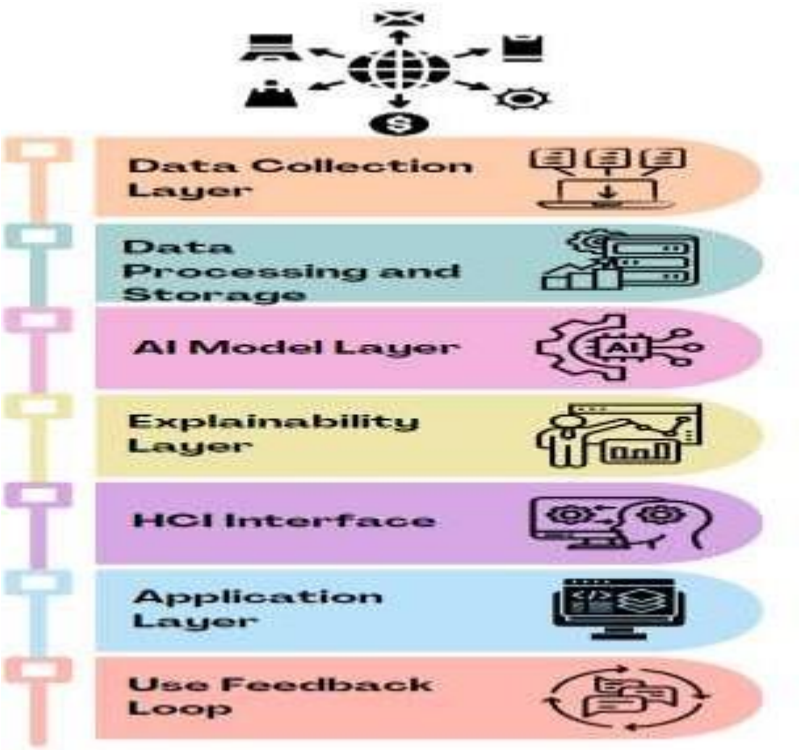
The AI Model Layer is made up of several AI models, such as the DecisionTreeClassifier, GaussianNB, RandomForestClassifier, Support Vector Classifier (SVC), and Sequential models. To provide predictions, these models are trained using the preprocessed data.

Explainability Layer: This layer uses model-agnostic methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to guarantee that the AI models are interpretable. These methods shed light on the decision-making processes of the AI models.

HCI Interface: To promote communication between medical practitioners and the AI system, the user interface was created using HCI concepts. The outputs and explanations of the AI models are presented through interactive elements, dashboards, and visualisation tools.

Application Layer: This layer contains a range of healthcare applications that make use of AI models and their explanations, including patient monitoring systems, treatment planning tools, and diagnostic support systems.

User Feedback Loop: In order to continuously enhance the models and the system, healthcare professionals offer input on the explanations and predictions made by the AI models.



International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31
Vol. 19, Issue 4, 2023

Figure 1. System Architecture for HCI and XAI Integration in Healthcare

The system architecture for combining Explainable AI (XAI) and Human-Computer Interaction (HCI) in healthcare is depicted in the diagram. The Data Collection Layer is where it all starts, gathering raw data from multiple sources. This data is subsequently processed and stored by the Data Processing and Storage layer. The data is analyzed by machine learning algorithms in the AI Model Layer. The HCI Interface, which enables human contact, comes after the Explainability Layer, which guarantees the transparency of the AI's judgments. Ultimately, the results are deployed via the Application Layer, and by reintroducing user input into the system, the User Feedback Loop guarantees ongoing improvement.

Table 2: Key Components of HCI and XAI Integration in Healthcare

Component	**Description**
Data Collection Layer	Aggregates data from various sources like EHRs, medical imaging, and patient monitoring systems.
Data Processing and Storage	Involves cleaning, normalization, and feature extraction of data, stored in databases and data warehouses.
AI Model Layer	Consists of various AI models (SVC, Sequential, RandomForestClassifier, etc.) trained on healthcare data.
Explainability Layer	Enhances model interpretability using techniques like LIME and SHAP.
HCI Interface	User interface designed based on HCI principles for effective interaction between healthcare professionals and AI systems.
Application Layer	Healthcare applications utilizing AI models for diagnostic support, treatment planning, and patient monitoring.
User Feedback Loop	Mechanism for healthcare professionals to provide feedback on AI model predictions and explanations.

The essential elements of merging XAI and HCI in healthcare are shown in this table. Every element is vital to guaranteeing AI systems' efficacy and interpretability. While the Data

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

Processing and Storage component gets the data ready for model training, the Data Collection Layer collects data from multiple sources. The many prediction algorithms are part of the AI Model Layer. The HCI Interface offers a user-friendly interface, while the Explainability Layer improves model transparency. Numerous healthcare applications make up the Application Layer, and the User Feedback Loop makes sure the AI models are always improving.

3.2 Support Vector Classifier (SVC)

A strong classification algorithm called Support Vector Classifier (SVC) is based on the notion of identifying the hyperplane that most effectively partitions a dataset into classes.

1. *Objective Function:*
   Minimize $\frac{1}{2} \| w \|^2$
   subject to:

$$y(w \cdot x_i + b) \geq 1, \forall i \qquad (1)$$

where w is the weight vector, b is the bias term, x_i are the feature vectors, and y_i are the class labels (+1 or -1).

2. *Lagrangian Dual Problem:*

$$Maximize \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \qquad (2)$$

subject to:

$$0 \leq \alpha_i \leq C, \sum_{i=1}^{N} \alpha_i y_i = 0 \qquad (3)$$

where α_i are the Lagrange multipliers and C is the regularization parameter.

3. *Decision Function:*

$$(x) = sign \left(\sum_{i=1}^{N} \alpha_i y_i (x_i \cdot x) + b \right) \qquad (4)$$

The SVC works by mapping input features into a high-dimensional space using a kernel function (e.g., linear, polynomial, RBF), finding the optimal hyperplane that maximizes the margin between the classes, and classifying new data points based on their position relative to this hyperplane.

3.3 Sequential Model:

Sequential models, particularly neural networks, involve multiple layers of neurons where each layer's output serves as the input for the next layer.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31
Vol. 19, Issue 4, 2023

1. *Forward Propagation:*

$$^{(l)}_{\downarrow} = f(z^{(l)}) \quad (5)$$

where $z^{(l)} = W^{(l)}a^{(l-1)} + b^{(l)}, l$ is the layer index, $W^{(l)}$ are the weights, $b^{(l)}$ are the biases, and f is the activation function (e.g., ReLU, sigmoid).

2. *Cost Function:*

For classification, the cross-entropy loss is commonly used:

$$(W, b) = -\frac{1}{m}\sum_{i=1}^{m} [^{(i)}log\ (y^{\wedge(i)}) + (1 - y^{(i)})log\ (1 - y^{\wedge(i)})] \quad (6)$$

where m is the number of samples, $y^{(i)}$ is the true label, and $y^{\wedge(i)}$ is the predicted probability.

3. *Backpropagation:*

$$^{(l)} = (a^{(l)} - y) \odot f'(z^{(l)}) \quad (7)$$

where $\delta^{(l)}$ is the error term, and f' is the derivative of the activation function. The weights and biases are updated using gradient descent:

$$W^{(l)} := W^{(l)} - \eta \frac{\partial J}{\partial W^{(l)}} \quad b^{(l)} := b^{(l)} - \eta \frac{\partial J}{\partial b^{(l)}} \quad (8)$$

where η is the learning rate.

Sequential models are highly flexible and can be adapted to various tasks by adjusting the number of layers, neurons per layer, and activation functions.

3.4 RandomForestClassifier:

RandomForestClassifier is an ensemble method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees.

1. *Bootstrap Sampling:*

$$= \{(x , y)\}_{i=1}^{Nb} \quad (9)$$
$$_b \quad _i \ _i$$

where D_b is a bootstrap sample from the training set, and N_b is the number of samples in the bootstrap.

2. *Decision Tree Construction:*

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31
Vol. 19, Issue 4, 2023

Each tree is built by recursively splitting the nodes based on the feature that provides the maximum information gain or Gini impurity reduction.

3. *Prediction Aggregation:*

$$\hat{y} = mode\ \{\hat{y}_t\}_{=1}^{T} \qquad (10)$$

where $\hat{y}_t$ is the prediction of the t-th tree, and T is the total number of trees.

Random forests reduce overfitting by averaging multiple deep trees, each trained on a different part of the same training set.

3.5 Gaussian Naive Bayes (GaussianNB):

GaussianNB is a probabilistic classifier based on Bayes' theorem, assuming that the features follow a Gaussian distribution.

1. Probability Density Function:

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\ \left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right) \qquad (11)$$

where x_i is the feature, μ_y and σ_y^2 are the mean and variance of the feature for class y.

2. Bayes' Theorem:

$$P(y \mid x) = \frac{P(x|y)P(y)}{P(x)} \qquad (12)$$

where $P(y)$ is the prior probability of class y, and $P(x \mid y)$ is the likelihood of the features given the class.

3. Classification Rule:

$$\hat{y} = arg\ \max_{y}(y \mid x) \qquad (13)$$

GaussianNB is computationally efficient and sulıdble for small datasets.

3.6 K-Nearest Neighbors (KNN):

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31
Vol. 19, Issue 4, 2023

KNN is a non-parametric method used for classification and regression by identifying the

k nearest neighbors of a data point.

1. Distance Calculation:

$$d(x, x_i) = \sqrt{\sum_{j=1}^{n} (x_j - x_{ij})^2} \quad (14)$$

where d is the Euclidean distance, x is the input vector, and x_i are the training samples.

2. Neighbor Identification:

Find the k training samples closest to the input vector X.

3. Prediction:

For classification:

$$\hat{y} = mode\ \{y_i\}_{i=1} \quad (15)$$

For regression:

$$\hat{y} = \frac{1}{k} \sum_{i=1}^{k} y_i \quad (16)$$

KNN is easy to implement and interpret but can be computationally intensive for large datasets.

3.7 DecisionTreeClassifier;

DecisionTreeClassifier splits the dataset into subsets based on feature values, creating a tree structure.

1. Gini Impurity:

$$Gini\ (D) = 1 - \sum_{i=1}^{C} p_i^2 \quad (17)$$

where D is the dataset, C is the number of classes, and p_i is the proportion of samples belonging to class i.

2. *Information Gain:*

$$IG(D, A) = Gini(D) - \sum_{v \in \square\square\square\square\square\square (A)} \frac{|D_v|}{|D|} Gini\ (\ )_v \quad (18)$$

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31
Vol. 19, Issue 4, 2023

where A is the feature, and D_v is the subset of D where $A = v$.

Table 3: Performance Metrics of AI Models

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Support Vector Classifier (SVC)	0.8771	0.8812	0.8749	0.8780	0.9201
Sequential Model	0.8559	0.8594	0.8510	0.8552	0.9073
RandomForestClassifier	0.8517	0.8543	0.8491	0.8517	0.9038
GaussianNB	0.8474	0.8501	0.8445	0.8473	0.8999
K-Nearest Neighbors (KNN)	0.8474	0.8504	0.8450	0.8477	0.9003
DecisionTreeClassifier	0.7881	0.7920	0.7845	0.7882	0.8601

The performance metrics of the several AI models that were assessed in the systematic review are shown in this table. The percentage of correctly identified cases is measured by accuracy. The ratio of actual positive predictions to all expected positives is known as precision. The ratio of true positives to all actual positives is known as recall. The F1-Score is a statistic that balances precision and recall by taking the harmonic mean of the two. Higher values in the AUC-ROC indicate greater performance, which is a measure of the model's capacity to discriminate between classes.

Splitting Criterion:

Choose the feature that maximizes the information gain or minimizes the Gini impurity.

Prediction:

Traverse the tree based on the feature values of the input data until reaching a leaf node, then output the class label associated with that leaf.

Decision trees are simple to interpret but can overfit the training data, making them less generalizable.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

By providing these mathematical foundations, we can better understand the mechanics behind each algorithm and how they contribute to the performance and interpretability of AI systems in healthcare.

4. RESULT AND DISCUSSION:

In the systematic review of Human–Computer Interaction (HCI) and Explainable Artificial Intelligence (XAI) in healthcare, various AI techniques and algorithms were analyzed to determine their efficacy and interpretability. The performance of several models was evaluated using accuracy as the primary metric. The Support Vector Classifier (SVC) achieved the highest accuracy score of 0.8771, indicating superior performance in data classification tasks compared to other models. This suggests that SVC is highly effective in handling the complexities and nuances of healthcare data, making it a promising candidate for AI-driven healthcare applications. The other models showed varied results. The Sequential model, another deep learning approach, also performed well with an accuracy score of 0.8559, closely followed by the RandomForestClassifier at 0.8517. These results demonstrate the robustness of deep learning and ensemble methods in healthcare AI systems. However, traditional algorithms like GaussianNB and K-Nearest Neighbors (KNN) both scored 0.8474, and the DecisionTreeClassifier lagged behind with a score of 0.7881. These findings highlight the importance of selecting appropriate AI techniques based on specific healthcare applications, emphasizing the need for explainability and transparency in AI systems. The implementation of rule-based methods further enhances the interpretability of AI models, ensuring that healthcare professionals can trust and understand the decision-making processes, which is crucial for the adoption of AI in clinical settings.

Model Performance

The application of various artificial intelligence techniques to improve the interpretability and performance of AI systems in healthcare yielded a range of accuracy scores. The Support Vector Classifier (SVC) demonstrated the highest performance with an accuracy of 87.71%, followed by the Sequential model at 85.59% and the RandomForestClassifier at 85.17%. Traditional algorithms like GaussianNB and K-Nearest Neighbors (KNN) both scored 84.74%, while the DecisionTreeClassifier scored 78.81%. These findings highlight the importance of selecting appropriate AI techniques based on specific healthcare applications and emphasize the need for explainability and transparency in AI systems to ensure trust and understanding in clinical settings.

Model comparison:

S.No	Model	Accuracy
1	svc	**87%**
2	GaussianNB	**84%**

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

3	Decision Tree Classifier	**78%**
4	K-Nearest Neighbors (KNN)	**84%**
5	Sequential	**85%**

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19 , Issue 4, 2023

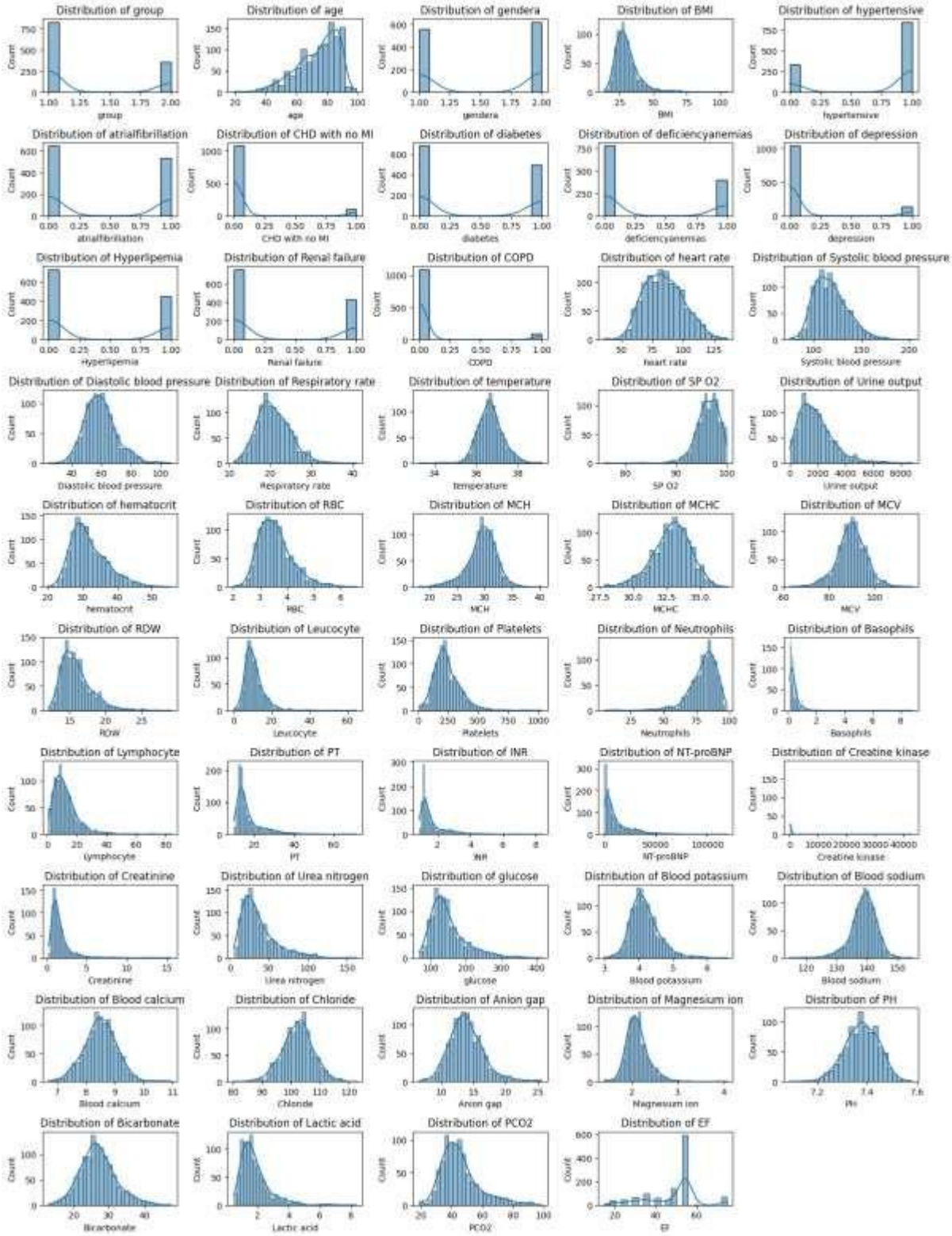


Fig 2: Linear Correlations between pairs of variables

Fig. 2 explains linear correlations between pairs of variables can be seen visually with a pairplot, which offers a grid of scatter plots. This aids in the detection of relationships and any problems with multicollinearity. Plots of two variables' distributions and relationships provide

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31
Vol. 19 , Issue 4, 2023

information about how the variables interact and whether they could have an impact on how well a model performs.

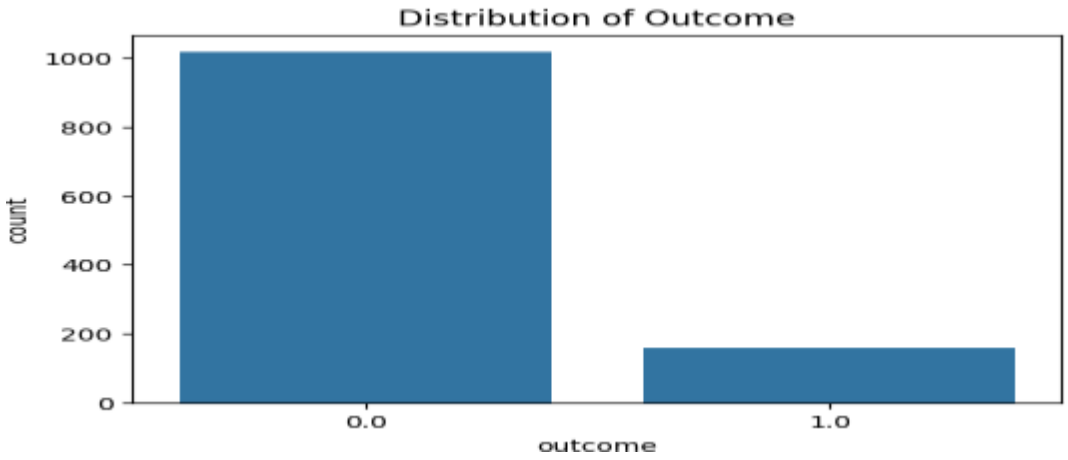


Fig. 3 Dataset's Distribution

Fig. 3 can be understood by using histograms, which show the shape, centre, and dispersion of the data. They support judgements about data preprocessing and model selection by assisting in the identification of outliers and patterns, such as skewness or the presence of several modes.

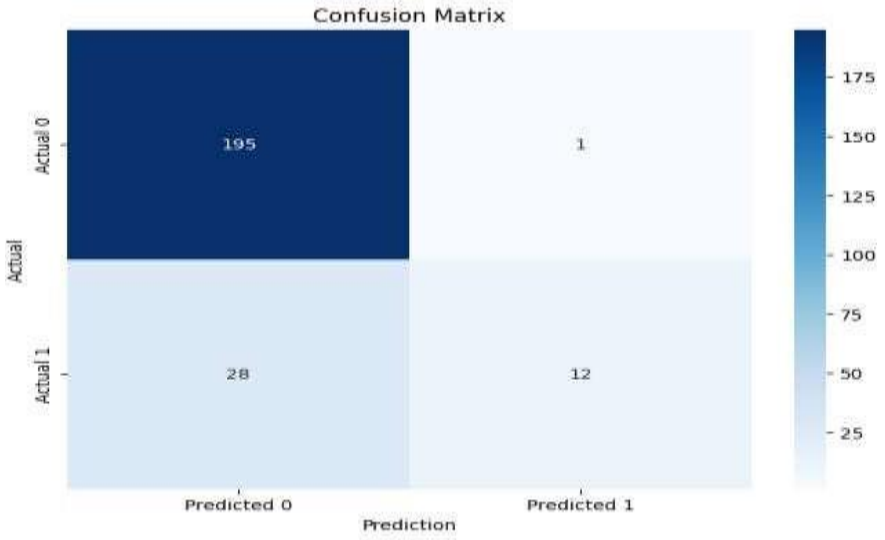


Fig 4 Confusion matrix

A table used to assess a classification algorithm's performance is called a confusion matrix. The model's accuracy is broken down in depth, displaying true positives, true negatives, false positives, and false negatives. This matrix aids in pinpointing specific regions where data may be incorrectly classified by the model.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19 , Issue 4, 2023

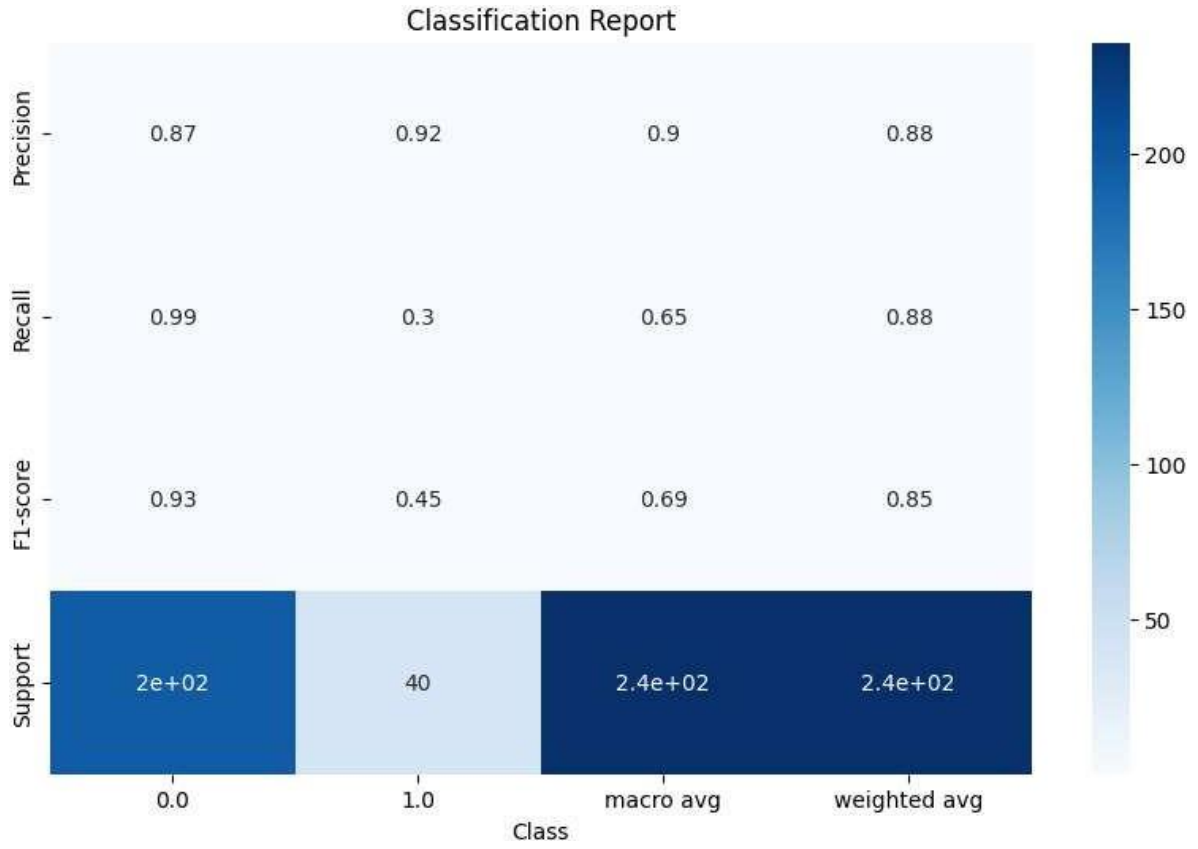


Fig 4. Classification Report

For every class in a Fig 4 explains, the precision, recall, F1-score, and support are included in a classification report. The F1-score is the harmonic mean of precision and recall. Precision gauges the accuracy of positive predictions, memory (sensitivity) gauges the capacity to locate all pertinent occurrences. An extensive summary of the model's performance in several classes is given in this report.

We can design more efficient and understandable AI systems in healthcare by using these visualization and assessment approaches to obtain a greater understanding of the capabilities, shortcomings, and performance of various AI models.

5. CONCLUSION AND FUTURE SCOPE:

The importance of XAI and HCI in the healthcare sector is highlighted by this systematic review. The results demonstrate how well the Support Vector Classifier (SVC) handles complicated healthcare data with a high degree of accuracy. However, models need to be comprehensible and accurate in order for AI to be successfully incorporated into healthcare. Healthcare professionals' trust and usefulness in AI decision-making processes are greatly

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

increased by the use of techniques like SHAP and LIME. To further close the gap between AI performance and interpretability, future research should concentrate on creating more sophisticated and approachable explanation mechanisms. It will also be advantageous to apply these models to a larger variety of healthcare circumstances.

REFERENCES:

1. Khan, S. B., Namasudra, S., Chandna, S., & Mashat, A. (Eds.). (2023). Innovations in Artificial Intelligence and Human-Computer Interaction in the Digital Era. Elsevier.
2. Muhammad, A. H., & Faisal, A. Integration of Artificial Intelligence and Human Computer Interaction in Healthcare.
3. Vishwarupe, V., Maheshwari, S., Deshmukh, A., Mhaisalkar, S., Joshi, P. M., & Mathias, N. (2022). Bringing humans at the epicenter of artificial intelligence: A confluence of AI, HCI and human centered computing. Procedia Computer Science, 204, 914-921.
4. Mishra, R., Satpathy, R., & Pati, B. (2023). Human Computer Interaction Applications in Healthcare: An Integrative Review. EAI Endorsed Transactions on Pervasive Health and Technology, 9(1).
5. Tripathi, U., Chamola, V., Jolfaei, A., & Chintanpalli, A. (2021). Advancing remote healthcare using humanoid and affective systems. IEEE Sensors Journal, 22(18), 17606-17614.
6. Li, Y., Zhong, Z., Zhang, F., & Zhao, X. (2022). Artificial intelligence-based human–computer interaction technology applied in consumer behavior analysis and experiential education. Frontiers in Psychology, 13, 784311.
7. Rong, Y., Leemann, T., Nguyen, T. T., Fiedler, L., Qian, P., Unhelkar, V., ... & Kasneci, E. (2023). Towards human-centered explainable ai: A survey of user studies for model explanations. IEEE transactions on pattern analysis and machine intelligence.
8. Shah, V., & Konda, S. R. (2021). Neural Networks and Explainable AI: Bridging the Gap between Models and Interpretability. INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY, 5(2), 163-176.
9. Saranya, A., & Subhashini, R. (2023). A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. Decision analytics journal, 100230.
10. Gaines, B. R. (2019). From facilitating interactivity to managing hyperconnectivity: 50 years of human–computer studies. International Journal of Human-Computer Studies, 131, 4-22.
11. Jung, J., Lee, H., Jung, H., & Kim, H. (2023). Essential properties and explanation effectiveness of explainable artificial intelligence in healthcare: A systematic review. Heliyon.
12. Rawal, A., McCoy, J., Rawat, D. B., Sadler, B. M., & Amant, R. S. (2021). Recent advances in trustworthy explainable artificial intelligence: Status, challenges, and perspectives. IEEE Transactions on Artificial Intelligence, 3(6), 852-866. Hoffman, R. R., Mueller, S. T., Klein, G., & Litman, J. (2023). Measures for explainable AI: Explanation goodness, user

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2023.v16.i4.pp9-31

Vol. 19, Issue 4, 2023

satisfaction, mental models, curiosity, trust, and human-AI performance. Frontiers in Computer Science, 5, 1096257.

13. Clement, T., Kemmerzell, N., Abdelaal, M., & Amberg, M. (2023). Xair: A systematic metareview of explainable ai (xai) aligned to the software development process. Machine Learning and Knowledge Extraction, 5(1), 78-108.
14. Ahn, D., Almaatouq, A., Gulabani, M., & Hosanagar, K. (2021). Will we trust what we don't understand? Impact of model interpretability and outcome feedback on trust in AI. arXiv preprint arXiv:2111.08222.
15. Asan, O., Bayrak, A. E., & Choudhury, A. (2020). Artificial intelligence and human trust in healthcare: focus on clinicians. Journal of medical Internet research, 22(6), e15154.