



Original Article

Explainable AI Models for Clinical Decision Support Systems

Chitiz Tayal

Senior Director, Data and AI.

Abstract - The data-driven world has led to the intervention of artificial intelligence and clinical decision support systems technologies in the healthcare sector in the decision-making process. However supportive these advanced systems are, they suffer from an issue of “black-box” or unexplainably of their working process, which hinders their utilisation in the health sector. Explainable Artificial Intelligence can mitigate such challenges by providing automation models that are transparent to healthcare individuals about their working procedures, which promises the clinical trust in healthcare systems and aids them in efficient decision-making. The evaluation of the data and information through statistical analysis has outlined that explainable artificial intelligence is the epitome of clinical data accuracy. In addition to this, the evaluation of the data through graphical representation has also outlined that, specifically, gender-based specialisation, the male patients mostly face problems due to respiratory conditions, along with heart conditions, while most of the female patients have faced diabetes. Explainable artificial intelligence has the capability in enhancing the data accuracy of the patients and developing treatments according to it.

Keywords - Clinical Decision Support Systems (CDSS), Artificial Intelligence (AI), Explainable Artificial Intelligence (XAI), Machine Learning (ML)

1. Introduction

1.1. Background

Clinical Decision Support Systems (CDSS) are being influenced by Machine Learning (ML) powered Artificial Intelligence (AI) methods, which assist healthcare professionals in the prediction of the probable outcome of patients. AI-based CDSS often result in minimal improvement since the course of action of the AI systems is undefined, and it becomes difficult to convince the healthcare professionals to consider the recommended system [1]. The lack of transparency among the AI models imposes challenges of interpretation of approaches behind their decisions, and it becomes harder to gain clinical trust [2]. Explainable AI models bridge the gap between complex algorithms and interpretability, which is essential for supporting clinical decisions.

1.2. Problem Statement

It is crucial to take into consideration that many AI models are referred to as “black-boxes” since the internal working processes of the model remain opaque to individuals. The lack of exposure of the functional algorithm of the AI models hinders the implementation of CDSS, as it is not trustworthy. Clinical experts avoid the assumption of risks to patient health based on modernisation. The benefits of explainable AI Models act as an interface between human beings and AI systems [3]. This means that AI systems are comprehensible to healthcare professionals.

1.3. Objective of the paper

Explain ability of AI systems holds a greater value in the Clinical Decision Support System (CDSS). Explaining the inner working processes helps to judge whether the AI models are reliable or not. The adoption of Explainable Artificial Intelligence (XAI) helps in addressing the challenges faced by AI models in terms of transparency and comprehensibility [4]. The field aims at the development of AI models that provide accurate predictions along with explainable AI algorithms for an efficient decision-making process. The paper aims at the objectification of shortcomings where the AI systems lack in retaining trustworthiness among the healthcare systems and how these problems can be addressed by Explainable AI (XAI).

1.4. Contributions

Clinical Decision Support System (CDSS) aids in providing assistance to the individuals responsible for patient care with efficient decisions that will be impactful for healthcare proceedings. The digital transformation in the public sector has resulted in the entry of AI systems in healthcare. Explainable AI or ML models are assisting CDSS in many aspects, like the generation of patterns from historical data and contributing to improving patient healthcare [5]. Explainable Artificial Intelligence (XAI) comes along with key factors like ensuring trustworthiness and identification of biases for generalisation of AI models, which is discussed later in this work.

2. Related Work

2.1. Overview of AI in Healthcare

AI has been growing its interest in healthcare for a decade, and healthcare systems are adopting machine learning approaches such as deep neural networks [6]. Efficient AI systems in healthcare institutions are divided into fragments and distributed among enormous datasets with the vision of boosting healthcare proceedings. By utilising the business intelligence (BI) tools, technicians lay the foundation of advanced automation models that would benefit the healthcare system. Advanced AI tools are instrumental in the early detection of critical abnormalities like cardiovascular diseases and cancer. The intervention of AI has led to the swift generation of patient reports. Despite being advantageous in various healthcare fields, AI approaches have not been welcomed in several health sectors due to their shortcomings.

2.2. Challenges in the adaptation of AI

An AI-based application utilises the historical records from the healthcare systems and analyses these datasets to propose an efficient automation architecture in order to promote better patient care. These datasets originate from the sensitive patient records, and these records are utilised for developing smart AI models, which will optimise the traditional working procedure of the healthcare systems. These applications impose a threat to data privacy and the security of sensitive health care data [7]. In major instances, AI models are unexplainable about their decision-making process, which is a major obstacle in the establishment of the Clinical Data Support System (CDSS). Lack of interpretable AI models can lead to an inclination towards traditional processes of patient care, hindering the advancement of AI in the healthcare system.

2.3. Conceptualisation of Explainable AI (XAI)

It is essential for the advancement of clinical decision-making processes in the evolving healthcare landscape. One of the transformation technologies includes Explainable Artificial Intelligence (XAI), which is responsible for creating a conjunction between individuals and AI systems. XAI represents a paradigm transformation that offers insights to clinicians about the working process of the AI models and fosters an understanding of its effect on patient health [2].

2.4. XAI in Clinical Decision Support Systems

The advancement of big data analytics and AI has led to the implementation of these techniques in CDSS. AI-based CDSSs are instrumental in simplifying diagnosis, reports and cure by AI models. However, the black-box aspect of AI-based models hinders the intervention of CDSS, as there is hardly any insight available about the algorithm upon which the model is acting. Explainable AI, or XAI, proposes ML techniques that are responsible for developing more explainable models along with sustaining higher performance. XAI-based CDSS is capable of responding to the decision-making process and benefits healthcare professionals in providing solutions to real-world problems [8].

2.5. Literature Gap

The AI-based CDSS has certain limitations that require future research. The majority of the available AI-based CDSS are dependent only on text data. If the data usage range is extended over intervals, only one additional type is used. One of the major issues faced by these advanced systems is that the cost of maintenance and installation is too high to be affordable for smaller healthcare bodies [8]. These challenges prevent the effective utilisation of these advanced technologies.

3. Methodology

3.1. Data source

The methodology system directs the interpretation of the research importance and its effects on the findings [9]. To learn how explainable AI can be relevant to clinical decision support systems 1000 entries of patients were analysed. The integration of such entries proved efficient in acquiring appropriate insight into the purpose of explainable AI in accruing CDSS by assessing the information of the patients and devising the solutions to improve their treatments. This was a highly accurate and effective methodological valuation in the identification of the models that were associated with this emerging technology and its use in facilitating optimal operation of CDSS.

3.2. Method of analysis

The analysis technique that will be used in this study is the use of logistic regression and a support vector machine. The SVMs stand as the abbreviation of support vector machines, which is the supervised variant of a machine learning algorithm, resulting in the categorization of the data that is able to maximise the prediction of the data variables [10]. This fusion of models was correct in achieving an understanding of the explains AI manifestation in improving the competence of CDSS. The rationality of SVM consists of the accretion of implicit regulations and the control of the linear and non-linear data through clarity, and then enhancing the efficiency of the data and information.

3.3. Ethical consideration

Ethical consideration is the epitome of maintaining the morality of the research through proper validation of the resources [11]. The anonymity of the healthcare professionals was maintained during the time of collecting data. In addition to this, the collected data was securely stored in the system of the researcher with a valid password and antivirus software. The sole

purpose of using the responses is to gain a strong understanding of how XAI have a strong influence within the various operations of healthcare. Further explanation illustrated that the data and information are kept confidential under the General Data Protection Regulations. Lastly, post completion of the research, the data and information are going to be de-identified.

4. Analysis of the Results and Discussion

4.1. Statistical analysis

4.1.1. Descriptive Statistics

Descriptive statistics are known as the summarisation of the data by describing its main features through the analysis and evaluation of the data set over a long period.

Table 1. Descriptive Statistics

Elements	Age	Gender	Blood_Pressure	Heart_Rate	Glucose_Level	Cholesterol	Body_Temperature	WBC_Count	Oxygen_Saturation	BMI	Model_Confidence	Feature_Importance_Score	Explanation_Clarity_Score	Local_Interpretability	Global_Transparency_Score	Disease_Prediction	Clinical_Domain
Mean	54.2	1.5	130.2	80.6	108.7	200.5	37.0	8.0	96.0	26.8	0.8	0.6	0.8	0.7	0.7	0.5	3.5
Standard Error	0.6	0.0	0.5	0.3	0.9	1.2	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1
Median	54.0	1.0	130.7	80.9	109.0	199.2	37.0	8.0	96.1	26.9	0.8	0.7	0.8	0.7	0.7	1.0	4.0
Mode	81.0	1.0	131.3	82.3	91.3	194.9	36.6	7.5	95.5	27.4	1.0	0.7	0.6	0.6	0.8	1.0	4.0
Standard Deviation	20.5	0.5	14.6	10.0	30.0	39.2	0.7	2.0	2.0	4.0	0.1	0.2	0.1	0.2	0.2	0.5	1.7
Sample Variance	421.1	0.3	214.5	100.9	899.6	1536.5	0.5	3.9	4.2	15.7	0.0	0.0	0.0	0.0	0.0	0.3	2.9
Kurtosis	-1.2	-2.0	-0.1	0.0	-0.2	0.3	-0.1	0.2	0.1	-0.2	-1.2	-1.2	-1.2	-1.2	-1.2	-2.0	-1.2
Skewness	0.0	0.0	-0.1	-0.1	0.0	0.2	0.0	-0.1	-0.1	0.0	-0.1	0.0	0.0	0.1	0.0	0.0	0.0
Range	69.0	1.0	90.8	68.9	187.8	283.9	4.2	14.2	12.8	25.7	0.5	0.7	0.5	0.6	0.6	1.0	5.0
Minimum	20.0	1.0	79.7	46.1	9.8	78.2	34.7	0.6	89.1	14.1	0.5	0.3	0.5	0.4	0.4	0.0	1.0
Maximum	89.0	2.0	170.5	115.0	197.6	362.1	38.9	14.8	101.9	39.8	1.0	1.0	1.0	1.0	1.0	1.0	6.0
Sum	54195.0	1498.0	130168.3	80553.3	108686.0	200483.2	37023.5	7969.7	95998.0	26791.3	753.6	649.6	750.9	701.2	699.2	504.0	3511.0
Count	100.0	10.0	1000.0	1000.0	1000.0	1000.0	1000.0	1000.0	1000.0	100.0	1000.0	1000.0	1000.0	1000.0	1000.0	1000.0	1000.0
Confidence Level(95.0 %)	1.3	0.0	0.9	0.6	1.9	2.4	0.0	0.1	0.1	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.1

The analysis of the table has revealed that the average age of the patients is 54.2, with moderate variability as calculated from the standard deviation, has amounts to 20.5. Most of the vitals have mainly fallen under normal range, though among them, cholesterol has showcased a high standard deviation of 39.2. The mean value being 0.8 has also illustrated a strong model confidence, and the prediction of the diseases was balanced as well. A well distribution of the clinical domains was also observed among the 6 categories, and the data quality appears to have minimal skewness across most of the variables.

4.1.2. Correlation

Correlation is responsible for measuring the direction and strength of a linear relationship that was constructed between two variables, and this linear relationship always ranges between -1 to +1.

Table 2. Correlation

Elements	Age	Gender	Blood_Pressure	Heart_Rate	Glucose_Level	Cholesterol	Body_Temperature	WBC_Count	Oxygen_Saturation	BMI	Model_Confidence	Feature_Importance_Score	Explanation_Clarity_Score	Local_Interpretability	Global_Transparency_Score	Disease_Prediction	Clinical_Domain
Age	1																
Gender	0.023878522	1															
Blood_Pressure	0.065731266	0.014553487	1														
Heart_Rate	0.008652104	0.00554741	0.017103465	1													
Glucose_Level	0.047460824	0.020586382	0.015717125	0.00882134	1												
Cholesterol	0.023574782	0.031055967	0.02294558	0.022951962	0.016732741	1											
Body_Temperature	0.006705543	0.036830752	0.010663599	0.029087002	0.055051978	0.007879865	1										
WBC_Count	0.042802193	0.006603545	0.020839169	0.064259876	0.023494664	0.014029564	0.029913067	1									
Oxygen_Saturation	0.027287993	0.078738394	0.01158885	0.026518311	0.025004919	0.011607263	0.006868367	0.0485078	1								
BMI	0.006684258	0.028391448	0.025864628	0.01574929	0.035266573	0.045653687	0.038847103	0.001736513	0.023668458	1							
Model_Confidence	0.020162707	0.014697998	0.007597801	0.030114654	0.034285401	0.02876151	0.015592126	0.009967174	0.074790824	0.018331521	1						
Feature_Importance_Score	0.021901195	0.040524156	0.018450105	0.0237021	0.00660658	0.07912083149	0.0485038657	0.00266158	0.028222723	0.0485038657	0.028222723	1					
Explanation_Clarity_Score	0.018153976	0.01866671	0.004674609	0.020995675	0.055705949	0.034178103	0.0225103	0.019885109	0.027488021	0.019885109	0.014966669	-0.020767918	1				
Local_Interpretability	0.031954474	0.05526797	0.017004929	0.034783581	0.096419272	0.00800808	0.00606551	0.049212469	0.055203824	0.049212469	0.066792529	0.063269903	0.0443086	1			
Global_Transparency_Score	0.01927144	0.040820137	0.008022526	0.01138313	0.035708455	0.039508455	0.002522658	0.04290118	0.01852396	0.04290118	0.010348907	-0.038211833	0.003717308	0.010646091	1		
Disease_Prediction	0.005725756	0.007968189	0.04588773	0.026924776	0.049474102	0.012667166	0.0049373456	0.00528094	0.04057322	0.04057322	0.040411023	-0.040500653	0.038642535	0.002161476	-1.88109E-05	1	
Clinical_Domain	0.006726496	0.005250251	0.07723763	0.024484112	0.059793662	0.022788965	0.00671777	0.012148195	0.09175184	0.0094796599	0.022827653	0.037142454	0.005768472	-0.013286376	0.004155279	-	1

The table containing the evaluation of the correlation between the variables has predominantly reduced a weak relationship, suggesting that maximum of the clinical variables opt for independent operational excellence. In addition to this, the table has also illustrated a notable relationship through a moderate but positive correlation between interpretability metrics. Contrary to this, a weak correlation was observed between feature importance and glucose, and the overall evaluation explained that there was no strong correlation with the existing prediction of diseases. Therefore, it has been indicated that the models of explainable artificial intelligence depend on complicated interactions rather than dependencies on single variables.

4.1.3. ANOVA

Analysis of variance is the full form of ANOVA, responsible for testing the statistical significance between means encapsulated in more than three groups. An extreme significance was showcased through the results presented in the table between the clinical variables and their critical value exceeding 1.64. This indicates that there is a substantial variation between the barrier metrics related to the patients, confirming that such clinical features are the representation of distinct measurements having various Central tendencies across the population accumulated in the dataset.

Table 3. Anova

Anova: Single Factor						
SUMMARY						
Groups	Count	Sum	Average	Variance		
Age	1000	54195	54	421		
Gender	1000	1498	1	0		
Blood_Pressure	1000	130168	130	214		
Heart_Rate	1000	80553	81	101		
Glucose_Level	1000	108686	109	900		
Cholesterol	1000	200485	200	1537		
Body_Temperature	1000	37024	37	0		
WBC_Count	1000	7970	8	4		
Oxygen_Saturation	1000	95998	96	4		
BMI	1000	26791	27	16		
Model_Confidence	1000	754	1	0		
Feature_Importance_Score	1000	650	1	0		
Explanation_Clarity_Score	1000	751	1	0		
Local_Interpretability	1000	701	1	0		
Global_Transparency_Score	1000	699	1	0		
Disease_Prediction	1000	504	1	0		
Clinical_Domain	1000	3511	4	3		
ANOVA						
Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	56590656.51	16.00	3536916.03	18787.14	0.00	1.64
Within Groups	3197263.30	16983.00	188.26			
Total	59787919.82	16999.00				

4.1.4. Evaluation of the data through graphical representations

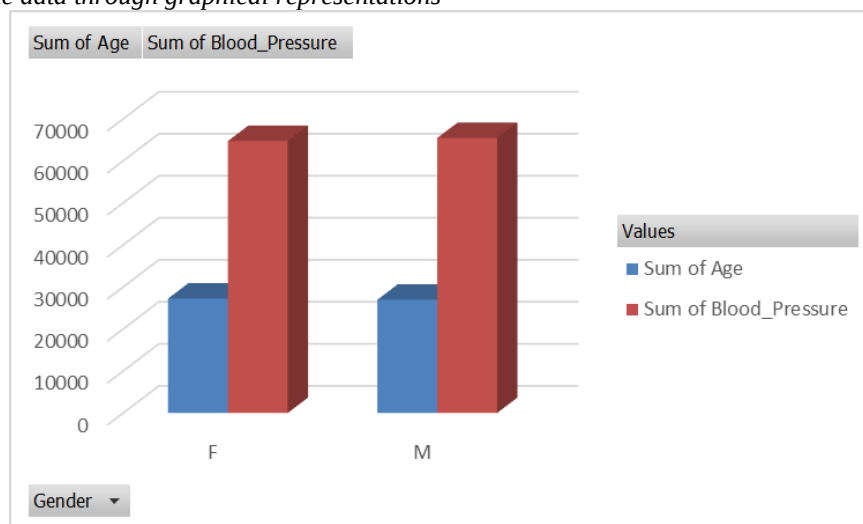


Figure 1. Blood Pressure as Per Gender and Age

The summation of the blood pressure of male and female individuals according to age is maximum for the male individuals when compared with the female individuals.

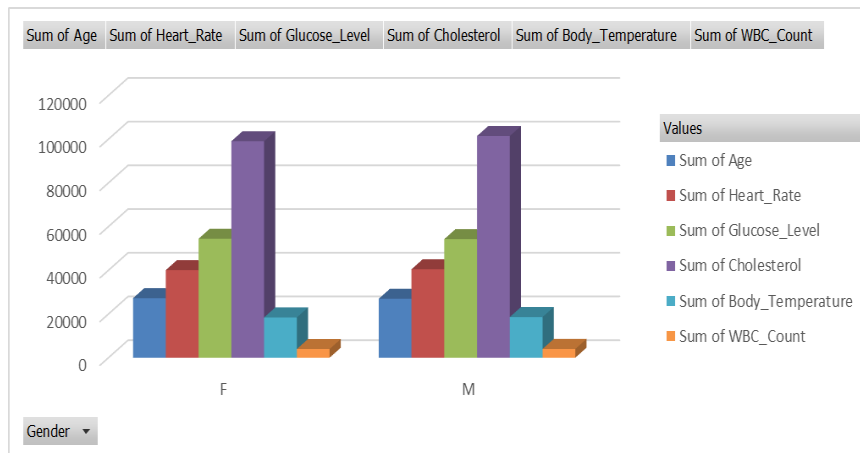


Figure 2. WBC Count, Temperature, Cholesterol, Glucose and Heart Rate Level as Per Age and Gender.

In a similar manner from the evaluation of the heart rate, glucose, cholesterol, temperature and WBC count, it has been observed that the increase of cholesterol is maximum for the male individuals.

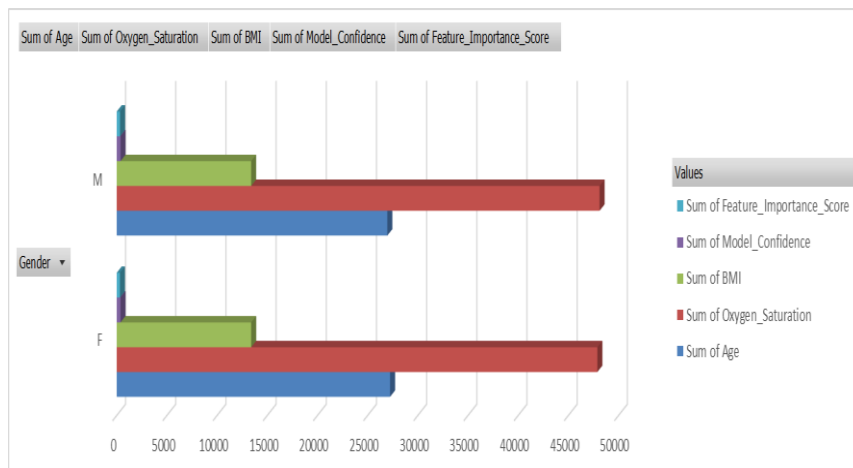


Figure 3. Oxygen Saturation, Model Confidence, and Feature Importance Score According to Gender and Age

A similar oxygen saturation was observed between the male and female patients, as shown in the above-mentioned figure.

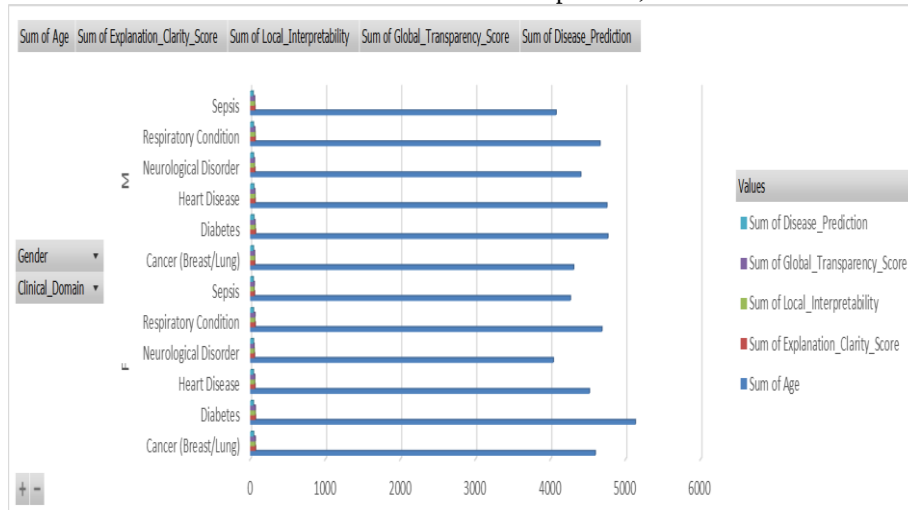


Figure 4. Explanation of the clarity score and interpretability as per gender and age

The explanation of the clarity and interpretability score as per the gender and age of the patients illustrated that the male patients are inclined towards diseases such as diabetes, heart disease and respiratory conditions, while female patients mostly face respiratory conditions and diabetes.

Therefore, the utilisation of the clinical decision support system through XAI is accurate in understanding the condition of the patients and generating better treatments as per the condition.

4.2. Evaluation of Support Vector Machine through performance metrics

From the evaluation of the dataset, the following are the important calculations:

- True Positive (TP): Actual is 1, predicted is 1 = $244 * (1000/632) = 386$
- False Positive (FP): Actual is 0, predicted is 1 = $75 * (1000/632) = 119$
- False Negative (FN): Actual is 1, predicted is 0 = $72 * (1000/632) = 114$
- True Negative (TN): Actual is 0, predicted is 0 = $241 * (1000/632) = 381$

4.2.1. Calculation of the performance metrics

Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy: $386+381/1000= 0.767$

Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision: $386/386+119= 0.7644$

Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall: $386/386+114= 0.772$

F1-Score

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

F1-Score= $2 * (0.7644*0.772)/(0.7644+0.772)= 0.768$

Therefore, the final evaluation of the SVM performance metrics for 1000 entries is described in the following table:

Table 4. Finalisation of the Performance Metrics of SVM

Metrics	Value
Accuracy	0.767
Precision	0.764
Recall	0.772
F1-Score	0.768

The consistent utilisation of the SVM model indicated that 76.7% is the percentage of all correct predictions. 76.4% outlined accurate but positive predictions; similarly, 77.2% is the amount of actual diseases detected. Finally, 76.8% is the accumulation of the balanced score between recall and precision.

4.3. Discussion

The performance evaluation of the statistical data demonstrated that explainable artificial intelligence has the capability of achieving predictive but balanced performance, along with the maintenance of interpretability within the clinical setting. The use of the model called support vector machine has a consistent accuracy of 76.7% and F1 score, during 76.8% aligned with the findings that explainable artificial intelligence has the capability of maintaining clinical accuracy by presenting transparent but accurate reasoning [12]. The weak correlation here has also outlined a limited relationship between the support of the clinical variables conceptualising the fact that medical diagnoses are strongly dependent on complicated but multi-factorial introduction dependencies on single variables.

The demographic patterns revealed through the graphical analysis particularly centralising on gender specific inclinations, outline the fact that explainable artificial intelligence is the epitome of presenting personalised support in developing accurate clinical decisions [13]. The substantial utilisation of variables through ANOVA has also validated the fact that clinical

diversity is important for robust generalisation of the model. Further evaluation of this information outlines the balanced performance of recall and precision, indicating the importance of fair construction of algorithms within the healthcare system through the utilisation of artificial intelligence, particularly in the area of equal detection rates in the area of patient sub-groups.

However, the evaluation of the data and information has also highlighted the fact that balancing interpretability followed by the power of prediction is very important for gaining accurate insight from the moderate performance of the dataset. The demonstration of the model support vector machine has outlined the importance of clinical utility and understanding the knee for expert accuracy in diagnosis for a better clinical decision support system. Therefore, the alternative advancement of artificial intelligence in the diagnostic system has the capability of strengthening not only the clinical expertise but also to present proper treatment Strategies for the patients.

5. Conclusion and Future Work

5.1. Conclusion

The AI-based solutions are more powerful in terms of solving the medical decision-making processes. However, they are very far in case of its application in real world. The implementation of AI systems in healthcare is conditioned by a number of factors. Axenicity and explicability of AI models are part of the requirements of encouraging clinical reliability [14]. Explainable AI and CDSS must work in order to guarantee the improvement of patient healthcare. Improvement of such intelligence systems in the health sector has the ability to introduce a global change in the mode of treatment.

5.2. Future work

Medical community is tasked with the role of introducing the optimal means of nurturing patient care [14]. Further researchers can point at several features of XAI-based CDSS in resolving the most important healthcare challenges such as abdominal surgeries and heart transplantations. This would be a new door of opportunities to the upcoming surgeons to perform a critical surgery in the coming days.

References

- [1] J. Amann *et al.*, "To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems," *PLOS Digital Health*, vol. 1, no. 2, p. e0000016, Feb. 2022, doi: <https://doi.org/10.1371/journal.pdig.0000016>.
- [2] N. Rane, S. Choudhary, and J. Rane, "Explainable Artificial Intelligence (XAI) in healthcare: Interpretable Models for Clinical Decision Support," *Social Science Research Network*, Nov. 15, 2023, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4637897
- [3] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson, "Explainable artificial intelligence: an analytical review," *WIREs Data Mining and Knowledge Discovery*, vol. 11, no. 5, Jul. 2021, doi: <https://doi.org/10.1002/widm.1424>.
- [4] V. Hassija *et al.*, "Interpreting Black-Box Models: a Review on Explainable Artificial Intelligence," *Cognitive Computation*, vol. 16, no. 1, pp. 45–74, Aug. 2023, doi: <https://doi.org/10.1007/s12559-023-10179-8>.
- [5] Y. Du, Anna Markella Antoniadis, C. McNestry, F. M. McAuliffe, and C. Mooney, "The Role of XAI in Advice-Taking from a Clinical Decision Support System: A Comparative User Study of Feature Contribution-Based and Example-Based Explanations," vol. 12, no. 20, pp. 10323–10323, Oct. 2022, doi: <https://doi.org/10.3390/app122010323>.
- [6] M. Sujana *et al.*, "Validation framework for the use of AI in healthcare: overview of the new British standard BS30440," *BMJ Health & Care Informatics*, vol. 30, no. 1, p. e100749, Jun. 2023, doi: <https://doi.org/10.1136/bmjhci-2023-100749>.
- [7] Y. Y. M. Aung, D. C. S. Wong, and D. S. W. Ting, "The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare," *British Medical Bulletin*, vol. 139, no. 1, pp. 4–15, Aug. 2021, doi: <https://doi.org/10.1093/bmb/ldab016>.
- [8] M. S. Reed *et al.*, "Evaluating Impact from research: a Methodological Framework," *Research Policy*, vol. 50, no. 4, 2021, Accessed: Oct. 25, 2025. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0048733320302225>
- [9] M. N. Laryeafio and O. C. Ogbewe, "Ethical Consideration dilemma: Systematic Review of Ethics in Qualitative Data Collection through Interviews," *Journal of Ethics in Entrepreneurship and Technology*, vol. 3, no. 2, pp. 94–110, Aug. 2023, doi: <https://doi.org/10.1108/JEET-09-2022-0014>.
- [10] J. Bajwa, U. Munir, A. Nori, and B. Williams, "Artificial intelligence in healthcare: Transforming the practice of medicine," *Future Healthcare Journal*, vol. 8, no. 2, pp. 188–194, Jul. 2021, doi: <https://doi.org/10.7861/fhj.2021-0095>.
- [11] R. L. Pierce, W. Van Biesen, D. Van Cauwenberge, J. Decruyenaere, and S. Sterckx, "Explainability in medicine in an era of AI-based clinical decision support systems," *Frontiers in Genetics*, vol. 13, Sep. 2022, doi: <https://doi.org/10.3389/fgene.2022.903600>.