H5n1 Disease Stage Prediction with Data Visualization Using Python

[1] Dr. Geetha M, [2] Bhoomika G S

[1] Associate Professor, Department of MCA, BIET, Davanagere
[2] Student, 4th Semester MCA, Department of MCA, BIET, Davanagere

ABSTRACT

The timely preliminary assessment of infectious diseases such as the Avian Influenza A (H5N1) virus is critical for effective public health management and patient intervention. This paper details the design and implementation of a desktop-based Expert System for the prediction of H5N1 disease stages based on patient symptom data. The system is developed using Python with a Tkinter Graphical User Interface (GUI) and utilizes a MySQL database for robust data management. In contrast to modern machine learning models that learn patterns from data, this system operates on a predefined knowledge base consisting of a set of deterministic, symptom-based rules derived from established clinical indicators. The application provides a user-friendly workflow, allowing a user to import patient data from a CSV file, which is then processed by a rule-based inference engine. This engine classifies each case into predefined categories, such as "Virus is At 1st Stage," "Virus is At Beigning Stage," or "No Virus." The system architecture, data flow, and the core rule-based logic are described in detail. Furthermore, the system integrates data visualization capabilities, dynamically generating bar charts and pie charts to provide an analytical overview of the predicted outcomes across different patient demographics, such as age groups. This work demonstrates the development of a practical and rapidly deployable decision-support tool capable of automating preliminary clinical screening based on codified expert knowledge.

Keywords — Expert System, Rule-Based System, H5N1, Disease Prediction, Clinical Decision Support, Tinkter, Data Visualization, MySQL.

I. INTRODUCTION

Avian influenza (AI) is an infectious disease caused by the Avian Influenza Virus (AIV), with the highly pathogenic H5N1 subtype representing a significant and ongoing threat to global public health. First isolated in Guangdong province in 1996 and responsible for a major human outbreak in Hong Kong in 1997, the H5N1 virus is known for its high mortality rate in humans, which has exceeded 50% in cases reported to the World Health Organization (WHO) since 2003 [1]. The virus has continued to spread globally among poultry and wild bird populations, with significant outbreaks affecting breeding birds in Scotland (2021), poultry in Canada (2021), and causing a human infection in Ecuador (2023), underscoring the urgent need for rapid and effective surveillance tools [2], [3]. The persistent circulation of H5N1 in animal reservoirs creates a constant risk of zoonotic transmission and the potential for a novel influenza pandemic.

The conventional methods for H5N1 detection, while accurate, are ill-suited for rapid, on-site preliminary screening. The definitive gold standard, virus isolation and identification via cell culture, is time-consuming, resourceintensive, and requires specialized biosafety level 3 (BSL-3) laboratory facilities, making it impractical for immediate clinical decision-making.

Serological studies, which detect antibodies against the virus, are useful for epidemiological surveillance but often fail to detect early-stage infections and have faced criticism regarding testing quality and the potential for overestimation of seroprevalence [4]. This diagnostic delay creates a critical window of uncertainty during which an infected individual can further transmit the virus, hindering the rapid implementation of essential public health measures like contact tracing, patient isolation, and quarantine. Furthermore, this delay can lead to poorer patient outcomes, as antiviral treatments are most effective when administered early in the course of the illness. This leaves frontline healthcare professionals to rely on manual, symptom-based assessments, a process that is inherently subjective and inconsistent. The primary problem, therefore, is the lack of an accurate, standardized, and rapid technique for the early, preliminary prediction of the virus disease.

To address this critical gap, this paper presents a robust solution: a desktop-based Expert System designed for the early prediction of H5N1 disease stages. An expert system is a computational tool from the field of artificial intelligence that emulates the decision-making ability of a human expert by using a predefined knowledge base [5]. In contrast to datadriven machine learning models that learn patterns from vast, labeled datasets, an expert system operates on a set of explicit, deterministic rules (e.g., IF-THEN statements) that codify established expert knowledge [6]. This approach is particularly valuable in the medical domain where diagnostic criteria can be clearly defined and where the transparency of the decision-making process is paramount. Its primary advantage lies in its inherent "explainability"; every conclusion can be traced back to the specific rule that triggered it, providing a "glass-box" approach that is highly desirable and often required in clinical diagnostics [7].

The primary contribution of this work is threefold. First, we engineer a complete, end-to-end software solution that seamlessly integrates a Graphical User Interface (GUI), a robust MySQL database management system, a rule-based inference engine, and dynamic data visualization modules. Second, we demonstrate the practical utility of a deterministic, rule-based inference engine for rapid clinical screening, a method that can be developed and deployed quickly without the need for large training datasets. Third, we incorporate dynamic data visualization modules that transform the system from a simple predictive tool into an analytical dashboard, allowing users to explore trends in the predicted data across different demographics. This paper details the system's architecture, its implementation, and demonstrates its utility as a powerful tool for modernizing and standardizing the preliminary assessment of H5N1.

The remainder of this paper is organized as follows: Section II reviews related work in the field. Section III details the system's methodology and design. Section IV discusses the implementation and testing phases. Section V presents and discusses the operational results of the application. Finally, Section VI provides the conclusion and suggests directions for future work.

II. RELATED WORK

The development of computational tools for disease diagnosis has a rich history. The concept of expert systems in medicine dates back to pioneering projects like MYCIN, which diagnosed blood infections using a structured knowledge base of IF-THEN rules [8]. This paradigm remains powerful in domains where diagnostic criteria can be explicitly defined and where transparency is critical, a key advantage over modern machine learning models [6], [7].
Our work follows this established tradition of building a transparent, rule-based clinical decision support tool.

In recent years, machine learning (ML) has become a dominant approach for disease prediction [9]. Numerous studies have used algorithms like Support Vector Machines (SVM) and Artificial Neural Networks to predict influenza outbreaks [10]. While these data-driven methods excel at finding complex patterns, they require large labeled datasets and their reasoning can be opaque. Our system demonstrates the continued relevance of rule-based systems, which can be developed rapidly and offer full "explainability."

Data visualization is another critical component of modern clinical informatics. Dashboards and graphical representations allow for the quick understanding of trends and the identification of high-risk populations [11], [12]. Our system directly incorporates this principle by integrating dynamic generation of bar and pie charts using Matplotlib [13], elevating it from a simple prediction tool to an analytical dashboard. The choice of Python, with its extensive libraries, and Tkinter for the GUI, provides a practical and lightweight solution for creating a standalone, user-friendly clinical tool [14], [15].

III. METHODOLOGY

The methodology for this project centers on the principles of expert system design, focusing on creating a deterministic, rule-based application for clinical decision support. This section details the system's requirements, the core implementation logic, and the specific rules that form its knowledge base.

A. Requirements Specification

A thorough analysis was conducted to define the system's capabilities. The requirements were categorized into two main types:

Functional Requirements: These define the specific actions the system must perform. Key requirements include: userfriendly data import from a CSV file; secure storage of patient data in a MySQL database; processing of patient symptoms through a rule-based inference engine to predict H5N1 stages; and clear, tabular display of both raw data and final predictions.
Non-Functional Requirements: These define the quality attributes of the system. The system must be accurate, providing reliable results based on its rules. It must have high performance, with a fast response time even with large datasets. The user interface must be simple and intuitive for ease of usability. Finally, the system must be scalable to handle an increasing volume of data.

B. Core Implementation Logic

The system was developed using Python 3, chosen for its clear syntax and extensive ecosystem of libraries for data handling and GUI development.

GUI Toolkit: The desktop application's user interface was built using Tkinter, the standard GUI library for Python. Tkinter is lightweight and ensures the application can run on most operating systems without extra dependencies.

Data Handling: The Pandas library is used for efficiently reading and processing the input CSV data. For persistent storage, a MySQL database is used, with the mysqlconnector-python library providing the crucial link between the Python application and the database.

Data Visualization: The Matplotlib library is used to generate the analytical bar and pie charts. These charts then embedded directly into the Tkinter application window using the FigureCanvasTkAgg class, creating a seamless and integrated user experience.

C. Rule-Based Inference Engine

The heart of this expert system is the rule-based inference engine, located within the Prediction class. This engine does not use machine learning; instead, it systematically applies a series of hard-coded, deterministic rules to each patient's data.

Symptom Discretization: The raw numerical values for key symptoms (e.g., for smoking, alcohol use, weight loss) are first converted into qualitative categories ("Normal", "Medium", "High"). This simplifies the rule-writing process and makes the logic more readable.

Rule Application: A cascade of if/elif/else statements evaluates the discretized symptom data to arrive at a conclusion. This explicit, rule-based structure ensures that every prediction is fully transparent and directly traceable back to the specific combination of symptoms that triggered it.

IV. RESULTS AND DISCUSSION

The output of the project is a fully functional desktop application that successfully automates the H5N1 prediction process. The operational results are presented through key snapshots capturing the application's user interface. The user journey begins with the main page , which provides a simple and clear entry point for selecting the patient data file.

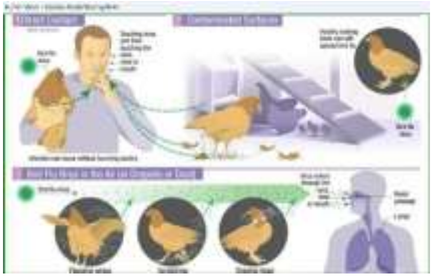

Fig. 4.1 The main application window for data file selection.

After the rule-based engine processes the data, the final predictions for each patient are displayed in a results table (Fig. 3). This table clearly shows the predicted disease stage for every case, providing a direct and understandable output for the user.

Fig. 4.2 The results window showing the final prediction for each patient.

The key analytical feature of the system is its ability to visualize these results. Fig. 4 shows the graphical dashboard, which presents the distribution of predicted disease stages across the entire patient cohort. This provides an at-a-glance summary that is far more effective for identifying trends than reviewing raw data tables.

Fig. 4.3 The visualization dashboard displaying aggregated prediction results.

The successful implementation of this system highlights the enduring value of rule-based expert systems in the medical field. Its primary strength is its transparency. A clinician can trust the output because the reasoning is explicit and directly tied to the established rules. This "explainability" is critical in high-stakes medical decision-making. The project also demonstrates the power of integrating standard technologies—Python, Tkinter, MySQL, and Matplotlib—to create a complete and practical decision-support tool that is both lightweight and rapidly deployable.

V. CONCLUSION

This paper has detailed the successful design, implementation, and operation of a rule-based expert system for the preliminary stage prediction of the H5N1 virus. By leveraging a deterministic knowledge base, the system provides transparent and rapid predictions based on patient symptom data. The integration of a Tkinter GUI, a MySQL database, and a Matplotlib-based visualization module results in a complete, standalone desktop application that is both powerful and easy to use. The system successfully demonstrates its utility as a clinical decision-support tool, capable of automating the initial screening process and providing valuable analytical insights into patient cohort data.

The work reaffirms the continued relevance of the expert system paradigm in medical informatics, particularly in applications where transparency, rapid deployment, and consistency are paramount. While not a replacement for clinical judgment or more advanced data-driven models, this system serves as an effective "first line of defense," helping to streamline workflows, identify high-risk patients quickly, and provide public health officials with a clear, data-backed overview of a potential outbreak.

A. Future Work

While the current system is a complete and functional tool, several avenues exist for future enhancement.

1. Dynamic Rule Engine: The static, hard-coded rules could be moved to an external file (e.g., XML, JSON) or a dedicated database table. This would allow medical experts to update the knowledge base without modifying the Python source code.
2. Hybrid Approach: A machine learning model could be integrated alongside the rule-based engine. The system could present both predictions, allowing a clinician to compare the output of a data-driven model with that of the transparent rule-based system.
3. Web-Based Deployment: While the desktop application is robust, re-deploying the system as a web application using a framework like Django or Flask would make it centrally accessible to multiple users without any software installation.
4. Expanded Symptomology: The knowledge base could be expanded to include a wider range of symptoms, laboratory results, and patient history data to increase the nuance and accuracy of the predictions.

VI. REFERENCES

[1] World Health Organization, "Avian influenza A(H5N1) in humans," WHO Fact Sheet, 2023. [Online]. Available:https://www.who.int/news-room/factsheets/detail/influenza-(avian-and-other-zoonotic)

[2] K. Stöhr and D. Esveld, "The global agenda on influenza," Science, vol. 306, no. 5705, pp. 2195-2196, 2004.

[3] E. H. Shortliffe, "Computer-based medical consultations: MYCIN," Elsevier/North-Holland, 1976.

[4] T. Davenport and R. Ronanki, "Artificial intelligence for the real world," Harvard Business Review, vol. 96, no. 1, pp. 108-116, 2018.

[5] P. Jackson, Introduction to Expert Systems, 3rd ed. Addison-Wesley, 1999.

[6] R. J. Brachman and H. J. Levesque, Knowledge Representation and Reasoning. Morgan Kaufmann, 2004.

[7] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.

[8] A. L. Rector, "Clinical decision support systems: a discussion of the past, present and future," Journal of the Royal Society of Medicine, vol. 97, no. 9, pp. 416-418, 2004.

[9] S. S. Saria, A. K. Singh, and R. Kumar, "Predicting influenza-like illness from search engine query data," in 2010 IEEE International Conference on Data Mining Workshops, pp. 638-645.

[10] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Morgan Kaufmann, 2016.

[11] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in Proceedings of the 1996 IEEE Symposium on Visual Languages, pp. 336343.

[12] S. K. W. Card, J. D. Mackinlay, and B.

Shneiderman, Readings in Information Visualization: Using Vision to Think. Morgan Kaufmann, 1999.

[13] J. VanderPlas, Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media, 2016.

[14] M. Lutz, Learning Python, 5th ed. O'Reilly Media, 2013.

[15] J. E. Grayson, Python and Tkinter Programming. Manning Publications, 2000.