Software Requirements Specification

Version 1.0
<<Annotated Version>>

MARCH 15, 2024

Parkinsons Disease Prediction Using ML Technique

Varun Gupta
Sumit Pal
Yash Jain

Index Page

5.1 Performance Requirements

5.2 Safety Requirements

5.3 Security Requirements

5.4 Software Quality Attributes

5.5 Business Rules

6. Other Requirements

- Scalability
- Interoperability
- Regulatory Compliance
- Localization

7. Appendices

- Appendix A: Glossary
- Appendix B: Analysis Models
- Appendix C: To Be Determined List

1. Introduction

1.1 Purpose

The purpose of this document is to outline the development and implementation of a Parkinson's disease prediction system using machine learning techniques. It serves as a guide for the design, development, and deployment of the system, with a focus on achieving high accuracy and precision in predicting the likelihood of Parkinson's disease. Additionally, the document aims to provide clarity on the scope, requirements, and constraints of the project, facilitating effective collaboration among stakeholders, including developers, healthcare professionals, and regulatory authorities. Ultimately, the purpose is to contribute to early diagnosis and intervention for Parkinson's disease, thereby improving patient outcomes and quality of life.

1.2 Document Conventions

Heading Styles:

1. Main headings are numbered and formatted in bold.

2. Subheadings are numbered and formatted in bold or italics.

Text Formatting:

- Key terms or important concepts are highlighted in *italics*.

- `Code snippets` or variable names are enclosed in backticks.

References:

References to external sources are provided in numbered format. Each reference includes the author(s), publication year, title, journal or website name, volume/issue (if applicable), page numbers (if applicable), and URL (if applicable).

Abbreviations and Acronyms:

- Abbreviations and acronyms are spelled out upon first mention, followed by the abbreviation/acronym in parentheses. Subsequent references use the abbreviation/acronym only.

Formatting Consistency:

- Consistent use of fonts, sizes, and styles throughout the document for readability and coherence.

- Adherence to any organizational or institutional style guidelines for consistency.

Version Control:

- Clear indication of document version and revision history to track changes and updates over time.

Language and Grammar:

- Use of clear and concise language to convey information effectively.

- Proper grammar, punctuation, and spelling to enhance readability and professionalism.

Visual Elements:

- Use of tables, diagrams, and charts to present information visually when appropriate.

- Consistent formatting and labeling of visual elements for clarity.

Accessibility Considerations:

- Ensuring that the document is accessible to users with disabilities, including screen reader compatibility and alternative text for images.

Review and Approval Process:

- Clearly defined process for document review and approval by relevant stakeholders to ensure accuracy and completeness.

This document is intended for developers, healthcare professionals, and stakeholders involved in the development and deployment of the Parkinson's disease prediction system.

1.4 Product Scope

The product scope encompasses the development of a Parkinson's disease prediction system, leveraging machine learning techniques for early detection and intervention. It aims to provide healthcare professionals with a robust tool to improve patient outcomes. The system will collect diverse patient data, preprocess it, and extract relevant features for model training. Various machine learning algorithms will be explored and evaluated for their efficacy in predicting Parkinson's disease, with a focus on achieving high accuracy and precision. Trained models will be deployed to predict disease likelihood in new patients, supported by interpretability features for clinician understanding. Seamless integration with existing healthcare systems will facilitate efficient data exchange and interoperability. A user-friendly web interface will allow easy interaction for healthcare professionals, ensuring accessibility for users with diverse technical backgrounds. Scalability will be ensured to accommodate growing datasets and user demand, supported by modular architecture and cloud-based infrastructure. Compliance with healthcare regulations and data security measures will be prioritized to safeguard patient privacy. Deliverables include a fully functional prediction system, comprehensive documentation, training materials, and ongoing maintenance and support services.

1.5 References

1	Parkinson's Disease Foundation. (n.d.). Understanding Parkinson's. Retrieved from https://www.parkinson.org/understanding-parkinsons
2	Parkinson's Disease Foundation. (n.d.). Understanding Parkinson's. Retrieved from https://www.parkinson.org/understanding-parkinsons
3	Tysnes, O. B., & Storstein, A. (2017). Epidemiology of Parkinson's disease. Journal of Neural Transmission, 124(8), 901–905.
4	European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L 119/1.
5	HealthIT.gov. (n.d.). Health IT Legislation and Regulations. Retrieved from https://www.healthit.gov/topic/health-it-legislation-and-regulations

2. Overall Description

2.1 Product Perspective

The Parkinson's disease prediction system stands as a pivotal addition to contemporary healthcare ecosystems, seamlessly integrating with existing platforms. By embedding within established healthcare systems and applications, it offers a streamlined pathway towards early diagnosis and intervention. This integration not only enhances the accessibility of predictive insights but also reinforces the collective effort towards combating Parkinson's disease.

2.2 Product Functions

Data Collection:

The system will orchestrate the gathering of diverse patient data encompassing demographic details, medical background, symptomatology, and potentially pivotal biomarkers like voice recordings or handwritten samples. This comprehensive approach ensures a holistic understanding of each patient's profile, empowering precise analysis and informed decision-making in Parkinson's disease diagnosis and management.

Preprocessing and Feature Extraction:

Before initiating model training, the accumulated data will undergo meticulous preprocessing stages, including normalization, imputation of missing values, and feature scaling. Leveraging sophisticated feature extraction techniques, the system will extract pertinent insights from raw data, whether it involves deriving statistical attributes from time-series data or implementing dimensionality reduction methodologies. This refined process ensures that the model is furnished with optimally processed and enriched data, thereby enhancing the accuracy and efficacy of Parkinson's disease prediction.

Model Training:

The system will embark on the pivotal stage of model training, where an array of machine learning algorithms including logistic regression, support vector machines, and sophisticated deep learning architectures will be employed. Harnessing the power of hyperparameter tuning and cross-validation techniques, the models will undergo rigorous optimization to attain peak performance and robust generalizability. This meticulous approach ensures that the system is armed with finely-tuned predictive capabilities, poised to deliver accurate and reliable predictions for Parkinson's disease prognosis.

Prediction:

Following the training phase, the models will possess the capability to forecast the probability of Parkinson's disease occurrence based on input data from new patients. Through the system's intuitive user interface, prediction results will be seamlessly presented to users, accompanied by informative confidence scores or probability estimates. This transparent and accessible manner of conveying predictions empowers clinicians and patients alike with valuable insights, facilitating informed decision-making and proactive management of Parkinson's disease.2.3 User Classes and Characteristics

Healthcare professionals

Data scientists

Patients

2.4 Operating Environment

Within the healthcare domain, the system will function seamlessly, leveraging access to patient records and pertinent medical data. Operating within this specialized environment ensures that the system is intricately woven into the fabric of clinical practice, enabling seamless integration with existing healthcare infrastructures. By harnessing the wealth of patient information available, the system is poised to enhance diagnostic precision and streamline decision-making processes in the management of Parkinson's disease.

2.5 Design and Implementation Constraints

Data Privacy Regulations:

The system must adhere rigorously to stringent data privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States or the General Data Protection Regulation (GDPR) in the European Union. Compliance entails robust measures for secure storage, transmission, and processing of patient data, coupled with the imperative of obtaining explicit consent from patients for data utilization.

Computational Resources:

Given the computational demands of training machine learning models, the system may necessitate substantial computational resources, ranging from high-performance computing clusters to cloud infrastructure. To alleviate resource constraints, strategies such as algorithm optimization and parallel processing techniques will be employed to streamline model training processes.

Integration:

Seamless integration with existing healthcare systems, electronic health records (EHR), or telemedicine platforms is paramount for the effective assimilation of the Parkinson's disease prediction system into clinical workflows. This necessitates strict adherence to interoperability standards and compatibility with diverse data formats and protocols, ensuring harmonious coexistence within the healthcare ecosystem.2.6 User Documentation

Comprehensive user documentation will be provided for system installation, usage, and maintenance.

2.7 Assumptions and Dependencies

Availability of Labeled Data:

The successful training of machine learning models relies on the availability of labeled data. It is assumed that an adequate amount of accurately labeled data for both Parkinson's disease and healthy individuals will be accessible for model training purposes.

Stable Internet Connection:

Dependence on a stable internet connection is assumed for seamless data transfer during model development and deployment phases. A consistent internet connection is also essential for real-time model deployment and integration into existing healthcare systems or applications.3. External Interface Requirements

3. External Interface Requirements

3.1 User Interfaces

The system will boast an intuitive web interface designed for effortless data input and seamless visualization of prediction results. This user-friendly interface ensures accessibility and ease of use for both healthcare professionals and patients alike.

3.2 Hardware Interfaces

Compatible with standard hardware configurations, the system seamlessly integrates with servers designated for model deployment, as well as end-user devices utilized for data input. This compatibility ensures flexibility and accessibility across various hardware setups.

3.3 Software Interfaces

Integration with a plethora of software components is pivotal for system functionality. This includes seamless interaction with databases for efficient data storage, integration with machine learning libraries for robust model training, and utilization of web frameworks for streamlined interface development. By interfacing with these essential software components, the system guarantees optimal performance and functionality.

3.4 Communications Interfaces

Facilitating seamless communication between the prediction system and external applications or databases, APIs play a pivotal role. These APIs enable efficient data exchange, ensuring interoperability and synergy between the prediction system and external entities. This streamlined communication enhances the system's versatility and utility within the broader healthcare ecosystem.4. System Features

Real-time Prediction:

The system will empower real-time prediction of Parkinson's disease likelihood, facilitating prompt intervention and strategic treatment planning for patients. Achieving this feat demands low-latency processing capabilities and the integration of efficient model inference algorithms. By ensuring swift and responsive performance, the system empowers healthcare professionals with timely insights, enhancing patient care outcomes.

Interpretability:

Beyond mere accuracy, the system prioritizes interpretability, offering comprehensive features to elucidate the reasoning behind each prediction. This entails highlighting pivotal features or variables that influence prediction outcomes, thereby fostering understanding and trust in the model's decisions among clinicians. By providing transparent insights, the system enhances clinical decision-making, ultimately optimizing patient care pathways.

4. System Features

4.1 System Feature 1: Real-time Prediction

Description:

This feature enables the system to deliver real-time predictions of Parkinson's disease likelihood based on input data. Upon receiving new patient data, the system swiftly processes the information and provides an immediate prediction, allowing for timely intervention and treatment planning.

Implementation Details:

- The system architecture incorporates low-latency processing mechanisms to ensure rapid response times.

- Efficient model inference algorithms are utilized to enable quick predictions without compromising accuracy.

- Real-time data streaming capabilities are integrated to facilitate continuous input and prediction generation.

Expected Outcome:

- Healthcare professionals can promptly assess the likelihood of Parkinson's disease in patients, enabling timely intervention and treatment planning.

- Patients receive immediate feedback on their condition, facilitating early engagement with healthcare providers and enhancing patient-centered care.

- The system's real-time prediction capability enhances overall efficiency in clinical decision-making and contributes to improved patient outcomes..

4.2 System Feature 2: Interpretability

Description:

This feature provides interpretability features to elucidate the rationale behind each prediction generated by the system. It aims to enhance understanding and trust in the model's decisions among healthcare professionals by explaining the important features or variables influencing the prediction outcome.

Implementation Details:

- The system employs explainable AI techniques, such as feature importance analysis and model-agnostic interpretability methods.

- Important features contributing to each prediction outcome are highlighted and presented to users in a comprehensible manner.

- Interactive visualizations may be utilized to facilitate exploration and understanding of prediction rationales.

Expected Outcome:

- Healthcare professionals gain insights into the factors influencing the prediction of Parkinson's disease likelihood, enabling informed decision-making.

- Increased transparency and interpretability foster trust in the predictive model, encouraging its adoption in clinical practice.

- Enhanced interpretability contributes to improved communication between healthcare professionals and patients, empowering shared decision-making processes.5. Other Nonfunctional Requirements

4.3 System Feature 3: Scalability

Description:

This feature ensures that the system can scale effectively to handle growing datasets and increasing user demand. It encompasses various scalability considerations such as horizontal scaling of computational resources, distributed data storage, and implementation of load balancing mechanisms.

Implementation Details:

- Horizontal scaling techniques are employed to expand computational resources horizontally, enabling the system to handle increased workload and processing demands.

- Distributed data storage solutions, such as NoSQL databases or distributed file systems, are utilized to accommodate growing datasets efficiently.

- Load balancing mechanisms are implemented to evenly distribute incoming requests across multiple servers, ensuring optimal utilization of resources and preventing bottlenecks.

Expected Outcome:

- The system can seamlessly accommodate the expansion of datasets and user base without compromising performance or responsiveness.

- Scalability measures enable the system to maintain high availability and reliability, even under heavy loads or spikes in user traffic.

- Scalability ensures that the system remains adaptable and future-proof, capable of meeting evolving requirements and scaling alongside organizational growth.

4.4 System Feature 4: Usability

Description:

This feature focuses on ensuring that the user interface of the system is intuitive and user-friendly, catering to users with varying levels of technical expertise. It encompasses usability testing and the implementation of user feedback mechanisms to continually enhance the user experience.

Implementation Details:

- The user interface is designed with a clean and intuitive layout, featuring clear navigation and well-defined functionalities.

- Usability testing sessions are conducted with representative users to identify usability issues and gather feedback for iterative improvements.

- User feedback mechanisms, such as surveys or feedback forms, are integrated into the system to collect user input and prioritize usability enhancements.

Expected Outcome:

- Users, including healthcare professionals, data scientists, and patients, can interact with the system confidently and efficiently, regardless of their technical background.

- Usability improvements based on user feedback lead to a more satisfying user experience, resulting in increased user adoption and satisfaction.

- The system's intuitive interface fosters productivity and reduces the learning curve for users, ultimately improving overall efficiency and effectiveness.

5. Other Nonfunctional Requirements

5.1 Performance Requirements

- Model Accuracy:The system must achieve a minimum accuracy of 90% in predicting Parkinson's disease to ensure reliable diagnostic outcomes.

- Response Time: Real-time prediction with minimal latency is crucial to enable prompt intervention and treatment planning.

5.2 Safety Requirements

Adherence to Data Privacy Regulations: Strict adherence to data privacy regulations such as HIPAA and GDPR to safeguard patient confidentiality.

- Error Handling: Robust error handling mechanisms must be in place to prevent inaccurate predictions and ensure patient safety.

5.3 Security Requirements

- Secure Data Transmission and Storage: Implement encryption protocols for secure data transmission and storage to mitigate the risk of unauthorized access or data breaches.

- Access Control: Enforce access control measures to protect sensitive patient information from unauthorized access.

5.4 Software Quality Attributes

- Robustness: The system should be robust enough to handle diverse input data and maintain stability under varying conditions, ensuring consistent performance.

- Maintainability: The system should be easy to update and modify as needed to accommodate evolving requirements and ensure long-term viability.

5.5 Business Rules

- Compliance with Healthcare Regulations: Strict adherence to healthcare regulations and ethical guidelines to ensure patient safety and regulatory compliance.

- Cost-Effective Implementation:Implement and maintain the system in a cost-effective manner to optimize resource allocation and maximize return on investment.

6. Other Requirements

- Scalability:The system should be designed to scale seamlessly to accommodate future growth in data volume and user base.

- Interoperability:Ensure compatibility with existing healthcare systems and standards to facilitate smooth integration and data exchange.

- Regulatory Compliance: Comply with relevant regulatory standards and certifications applicable to healthcare software, ensuring legal and ethical compliance.

- Localization:Support multiple languages and localization options to cater to diverse user populations and geographical regions.

7. Appendices

Appendix A: Glossary

- HIPAA: Health Insurance Portability and Accountability Act, a U.S. legislation ensuring the privacy and security of healthcare information.

- GDPR: General Data Protection Regulation, a European Union regulation governing data protection and privacy for individuals within the EU and the European Economic Area.

Appendix B: Analysis Models

System Architecture:

The system architecture will comprise several key components:

Data Ingestion Pipeline: Responsible for ingesting and preprocessing incoming data from various sources.

- Model Training Module: Executes the training of machine learning models using preprocessed data.

- Prediction Engine:Utilized for generating real-time predictions based on trained models.

- User Interface Layer: Facilitates interaction with end-users through a user-friendly interface.

Diagrams and flowcharts will be provided to illustrate the interaction between these components and the flow of data within the system.

Appendix C: To Be Determined List

This section will include any items or decisions pending further clarification or resolution, such as specific technology choices, integration requirements, or unresolved issues requiring further investigation.