CSCE 5214.004 SDAI

Group 9: DIABETES PREDICTION TOOL

Phase 5: Extending the architecture of an existing ML-based system

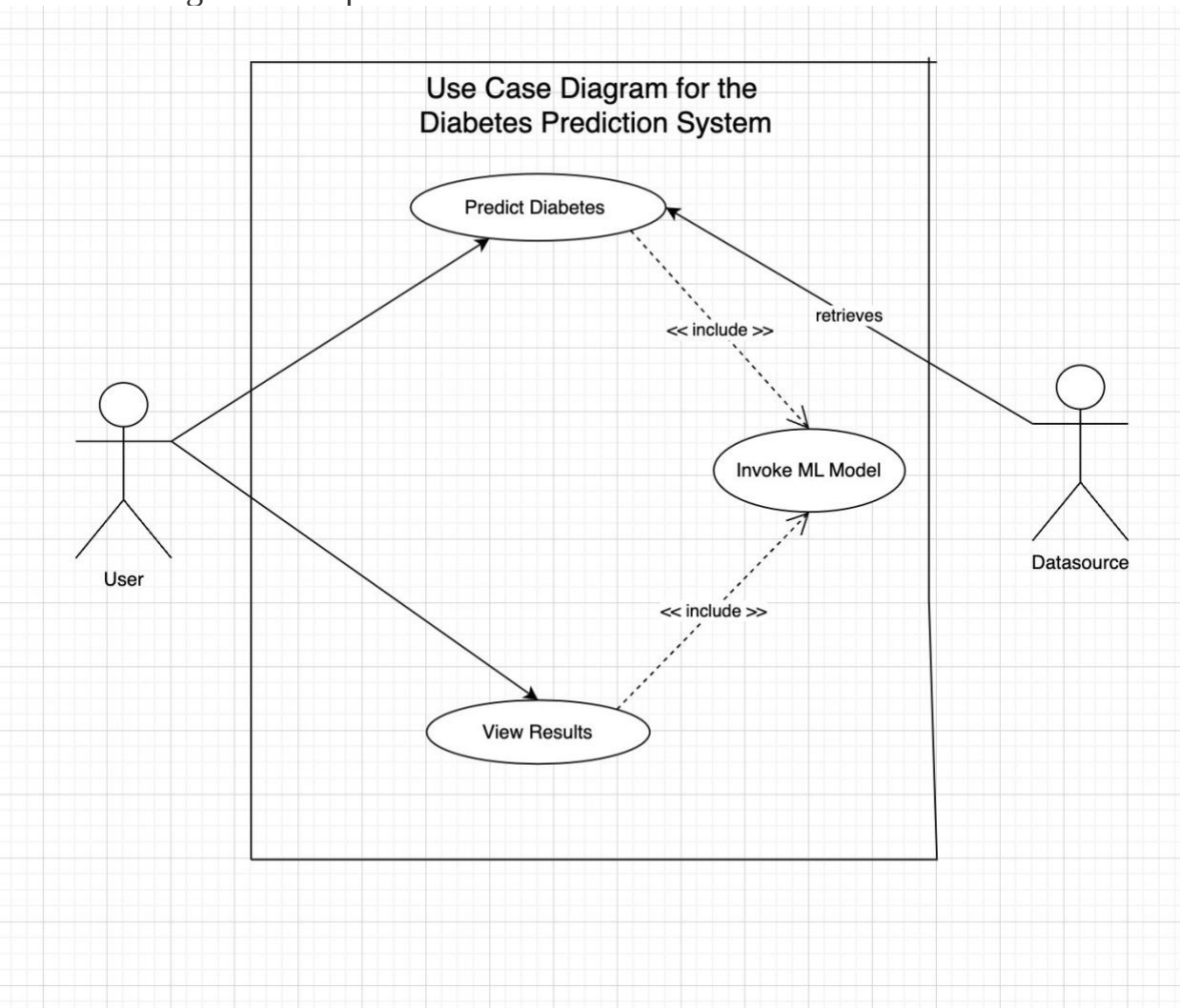
Aditya Vadrevu – 11601517
Suhas Siddarajgari Tellatakula - 11626111
Srikar Reddy Gunna - 11594012
Sai Praneeth Reddy Avula - 11582402
Indu Shashi Guda - 11618418

Task 1: Identification of architecture drivers

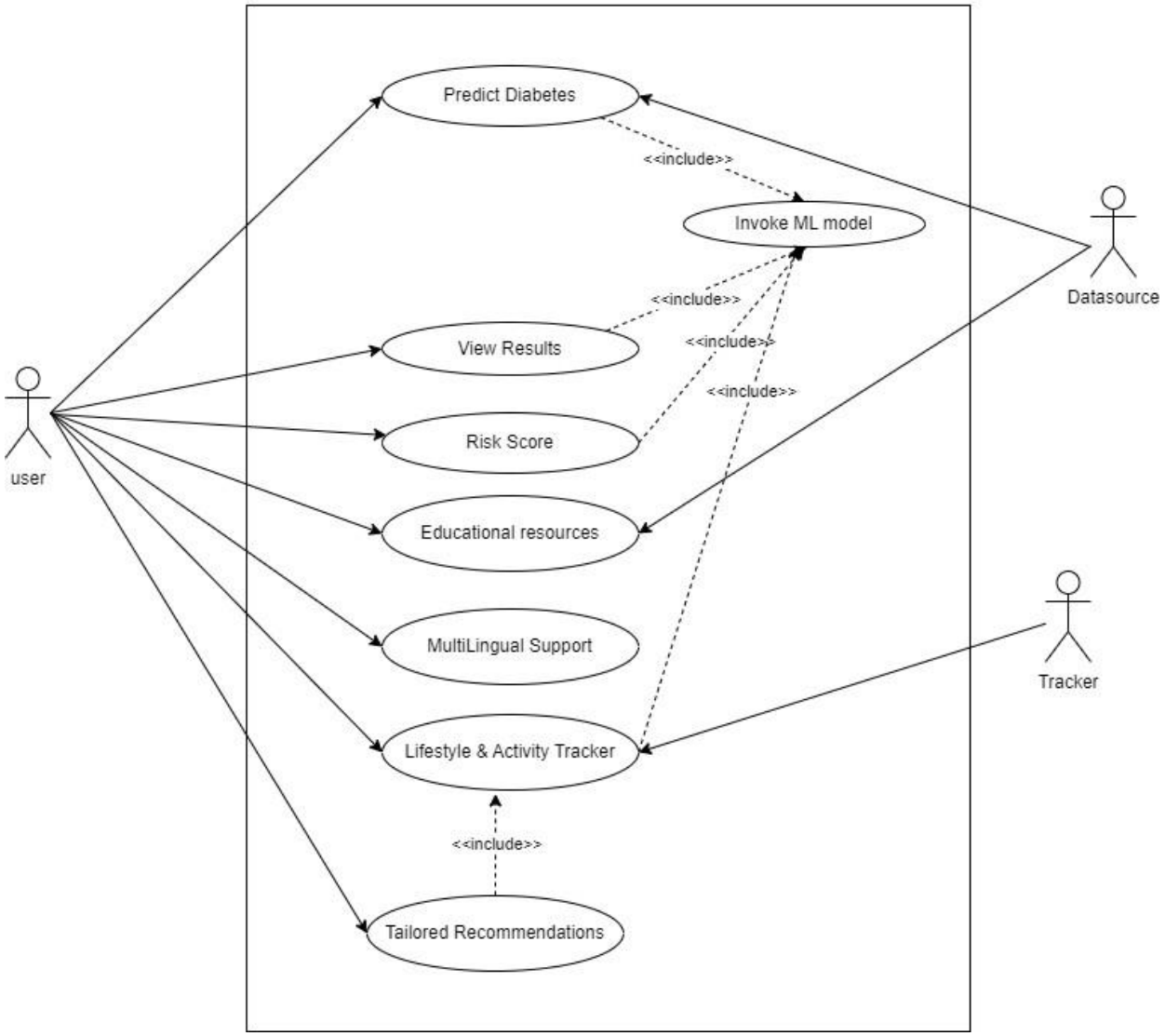
1.1. New Use Cases:

Use Case	Description
Risk Score	Rather than producing a straightforward binary forecast, the system might generate a continual diabetes risk score. Users would gain a more sophisticated picture of their personal risk as a result.
Tailored Recommendations	Using the user's risk score and other pertinent information, we can provide individualized suggestions for food adjustments, lifestyle adjustments, or preventive interventions. Customization along with Useful Advice.
Multilingual Support	Translate the tool into multiple languages to make it usable by a larger audience. This contributes to the system's increased accessibility and reach.
Educational Resources	Include instructional materials and informative data in the system for managing and preventing diabetes. This contributes to increasing Accessibility and Reach.
Lifestyle and Activity Tracking	Connect fitness trackers and other wearables to gather information on sleep habits, physical activity levels, and other lifestyle factors that affect diabetes risk. Data integration is a part of this.

Use case Diagram from phase 1



Updated Use Case Diagram



1.2. Quality Attributes: A table that specifies the identified quality attributes and associated use cases. You should report each of your quality attributes using the scenario format.

Use Case	Description
Risk Score	Performance: High throughput and low latency risk ratings should be provided by the system, particularly during periods of peak usage. Scalability: Both the quantity of data being processed and the number of users should be expanding within the system. Security: User data should be safely stored and safeguarded by the system, guaranteeing its integrity and confidentiality. Modifiability: The system must to be adaptable enough to take into account new risk variables and change as medical understanding does.
Tailored Recommendations	Performance: Personalized recommendations should be generated rapidly by the system using user data and risk assessments. Scalability: A high number of users with a variety of risk profiles should be supported by the system. Security: By safeguarding private information and limiting illegal access, the system should protect user privacy. Modifiability: New interventions, instructional materials, and lifestyle insights should all be able to be included by the recommendation engine.
Multilingual Support	Performance: With low latency and high accuracy, the system should be able to translate material and provide recommendations. Scalability: A large variety of languages and dialects should be supported by the system. Security: Regardless of the user's language, the system should protect user privacy and data security.

	Modifiability: It should be simple to add new translations and languages to the system.
Educational Resources	Performance: There should be less loading time and an efficient delivery of instructional content by the system. Scalability: A big amount of instructional content and a large number of users should be supported by the system. Security: The system needs to guarantee the validity and integrity of learning materials. Modifiability: Educational information should be simple to add, remove, and update via the system.
Lifestyle and Activity Tracking	Performance: Accurate and efficient data processing from a variety of wearables and trackers is required by the system. Scalability: The system must be able to manage massive amounts of data from numerous users and devices. Security: Information about a user's activities and way of life should be safely stored and protected by the system. Modifiability: The system must be able to change to accommodate new wearable gadget kinds and record more health information.

Additional Quality Attributes:

- Failure Detection: Failures in the system's many components—such as data processing, recommendation creation, and risk score calculation—should be identified and reported.
- Exception Handling: Exceptions should be gracefully caught and handled by the system to avoid cascade errors and system failures.
- Most Modified Components: The most frequent updates to include fresh information and resources are probably going to come from the recommendation engine and instructional materials.
- Data Frequency: The data source will determine how frequently data is obtained from outside sources. For instance, data on blood glucose may be acquired more frequently than information on lifestyle.

- Device Support: The estimated user base and usage habits will determine how many devices the system can support concurrently. Predicted demand and system resources should be taken into consideration when determining the precise amount.
- Reliability and Maintainability: Reduced downtime and data loss are desired, and the system should be extremely dependable and available. The system ought to be made as simple to upgrade and maintain as possible.
- Interoperability: It should be possible for the system to interface with other platforms and healthcare systems.

1.3. Constraints:

Use Case	Description
Risk Score	Users who rely too much on risk scores may experience anxiety or excessive tension. Along with risk assessments, the system should include concise justifications and background information.
Tailored Recommendations	Sophisticated algorithms and extensive user data are needed to produce recommendations that are tailored. Recommendations that are erroneous or irrelevant may result from inadequate data or excessively basic algorithms.
Multilingual Support	Language proficiency and familiarity with medical jargon are prerequisites for providing accurate and culturally relevant translations. Additionally, it can require a lot of resources to manage updates and content across several languages.
Educational Resources	Health and education specialists are needed to create and maintain high-quality educational content. Furthermore, it can be difficult to guarantee accessibility and modify material for a range of literacy levels.
Lifestyle and Activity Tracking	Compatibility issues and technological know-how are needed to integrate with a variety of wearable devices and measure different health metrics. Concerns about privacy and data accuracy must also be addressed.

Other Constraints:

- Data Availability and Quality: Obtaining complete and accurate user data is essential to producing trustworthy risk assessments and recommendations. Performance problems with the data, such as missing numbers or damaged data, can seriously affect the system.
- Privacy and Security Concerns: It is crucial to protect user data, especially private health information. To avoid unwanted access, data tampering, and loss, the system needs to adhere to data privacy laws and have strong security measures in place.
- Computational Resources: The computer power available for data processing, risk score computation, and suggestion generation determines the system's performance and scalability. Scalability, precision, and speed may need to be traded off due to resource constraints.
- Cost and Sustainability: There is a large financial commitment required for the system's development and maintenance. Maintaining operations and improvements over the long run requires a sustainable business model or funding source.

1.4. Concerns:

Concern	**Description**
Misinterpretation of Results	Misinterpretation of risk ratings or suggestions by users may result in unwarranted concern, self-diagnosis, or refusal to seek appropriate medical treatment.
Bias and Discrimination	The algorithms and data in the system might unintentionally reinforce preexisting prejudices in healthcare, which would have discriminatory effects on particular populations.
Overdependence on Technology	Users may get discouraged from seeking expert medical advice and interventions if they become overly dependent on the system for risk management and self-assessment.
Accessibility and Digital Divide	In order to use the system, not everyone may have access to the necessary technology or internet connectivity, which could exacerbate already-existing health inequities.

Overburdening Healthcare Systems	The system suggestions may result in a surge in testing and interventions, which could put a strain on healthcare practitioners and resources.

These limitations and worries emphasize how crucial it is to build, implement, and oversee the suggested system responsibly. Careful consideration of user needs, data privacy, ethical considerations, and potential biases will be necessary to address these challenges. To make sure the system helps with diabetes management and prevention, it will be essential to conduct ongoing evaluations and improvements.

Task 2: Update the software architecture

2.1. Solutions:

Solutions Addressing new use cases and quality attribute drivers

Category	Solutions
Quality Attribute Drivers	Performance: Optimize throughput while minimizing latency. Scalability: Able to manage a growing user base and volume of data. Security: Preserve user information, guarantee integrity and confidentiality. Adaptability to novel risk factors, therapies, and technological advancements. multilingual Support: Adapt to different accents and languages. Ensure that the educational resources are diversified and easily accessible. Way of Life and Exercise Tracking: Monitor a variety of statistics and integrate with wearables.
Architecture Patterns	Scalability and performance are increased using microservices architecture, which enables independent system component creation, deployment, and scaling. API Gateway: Simplifies integration and access control by acting as a single point of entry for external apps and devices. Event-driven architecture benefits from asynchronous communication and component decoupling, which leads to increased scalability and efficiency. Data Lake: Compiles and preserves data from multiple sources to support machine learning and data analysis for risk assessment and customized recommendations.

	Material management system: Facilitates the creation, translation, and updating of information while managing educational materials.
Architectural Tactics	Caching: Reduces latency and speeds up retrieval of frequently accessed material by storing it. Sensitive user data is protected both in transit and at rest by data encryption. Mechanisms for Access Control: Using user roles and permissions, limit access to information and features.. Localization is the process of modifying information and user interface for various language contexts. Standardized interfaces for data flow between system parts and external devices are defined by API standardization. Data quality checks: Before using data for risk assessment and recommendations, make sure it is accurate and full. maintaining privacy Methods: to preserve analytical insights while protecting user privacy, anonymize or aggregate data. Monitoring and Logging: Keeps tabs on user activity, system performance, and possible faults for troubleshooting and ongoing improvement.
Architectural Drivers	Performance: horizontal scaling, caching, and microservices architecture Scalability: data lakes, event-driven architecture, and microservices architecture Security measures include safe authentication, access control methods, and data encryption. Modifiability includes model versioning, rollback, API gateway, and microservices architecture. Support for Multiple Languages: Localization and Content Management System

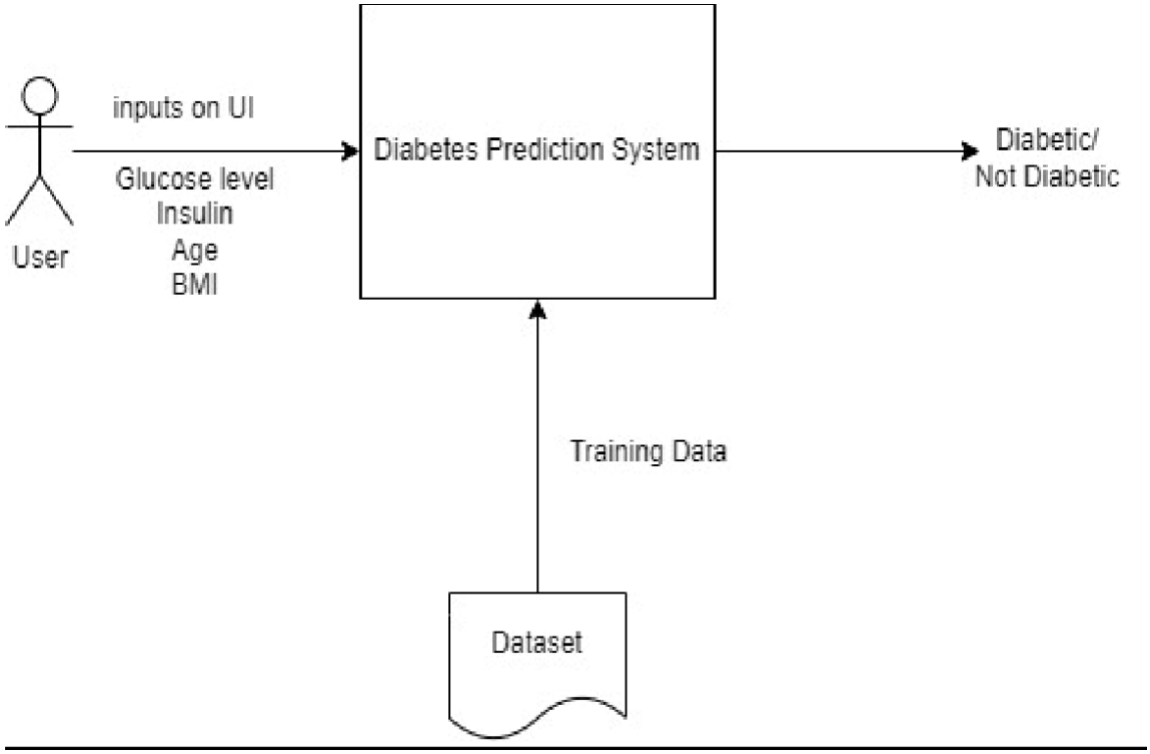
	Resources for Education: Content management system, features for accessibility Tracking lifestyle and activities: data quality checks and API standards.

Category	Solution
Data Availability and Quality	Work together with clinics, hospitals, and other healthcare providers to obtain anonymized patient data for the purpose of training and refining models. Establish pipelines for data cleansing and validation to find and fix biases, missing numbers, and inconsistencies. Through wearable integration, self-reporting, and approved data exchange with healthcare professionals, encourage users to give pertinent health data.
Privacy and Security	Implement robust data security measures, such as encryption, access control, and anonymization, to comply with regulations like HIPAA or GDPR. Provide clear and transparent information about data collection, usage, and storage practices, and obtain informed consent from users. Conduct periodic assessments of the system's security posture to identify and address potential vulnerabilities.
Computational Resources	To access scalable computational resources and facilitate economical operation, leverage cloud platforms. Use quantization and model compression strategies to lower resource needs and make mobile device deployment easier. To reduce storage costs and improve the speed of data access, make use of effective data storage solutions like data lakes and data warehouses.
Cost and Sustainability	Charge a subscription charge to get specialized recommendations or premium features. Investigate revenue-sharing or integration into current healthcare services through collaborations with healthcare groups. Seek financing for development and sustainability from governmental bodies, academic institutions, or charitable groups.
Risk Score Misinterpretation	Provide thorough justifications of risk ratings' implications, constraints, and suggestions for more action to go along with them. Include instructional resources in the system to describe diabetes, its risk factors, and its management techniques. Urge consumers to talk to their healthcare practitioners about their risk ratings to receive individualized advice and interpretation.
Bias and Discrimination	To prevent sustaining preexisting biases in the healthcare industry, train models using diverse and representative datasets. To find and fix any potential biases in the system's algorithms and data, do fairness tests and audits. To foster responsibility and trust, be open and honest about data gathering, model training techniques, and possible biases.
Overdependence on Technology	Encourage user education by informing users of the limitations of the system and stressing the value of consulting a medical expert for diagnosis and treatment. Establish a connection between the system and the current healthcare systems to

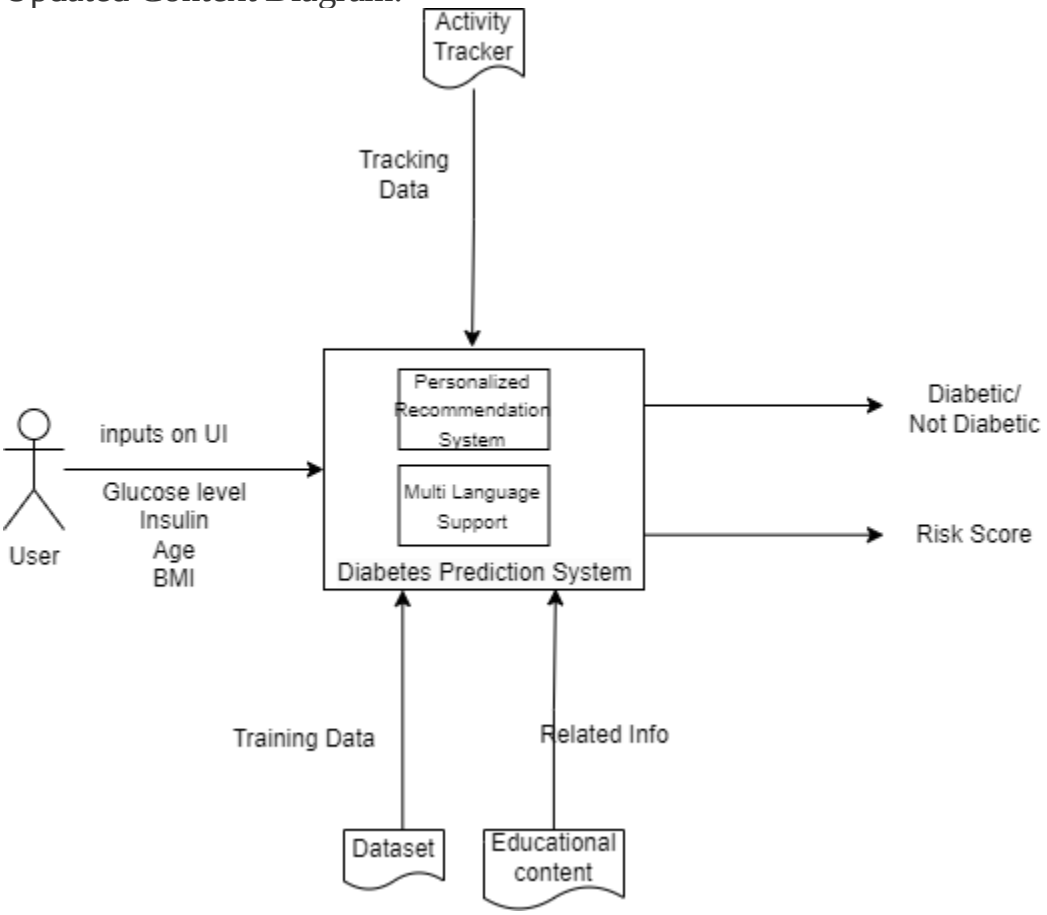
	facilitate easy access to medical personnel and resources. Promote proactive health management by encouraging people to take preventative actions and embrace healthy lifestyle practices.
Accessibility and Digital Divide	Create versions of your app for mobile and the web, and make sure it's accessible on a variety of platforms to accommodate people with varying tech access and tastes. Allow users with restricted internet connectivity to access critical functions and educational resources offline. Work together with neighborhood organizations to close the digital gap by giving people access to technology and training in digital literacy.
Overburdening Healthcare Systems	Concentrate intervention efforts and advice on users who have been identified as having significantly higher risk ratings. To enable users to independently manage their health, incorporate educational materials and self-management tools into the system. Collaborate with healthcare providers to incorporate the system into current processes and guarantee effective use of available resources.

2.2. Context Diagram

Context diagram from phase 1:



Updated Context Diagram:



Updates added are:

a) Risk score prediction system is added on top of the existing system to indicate continuous values.

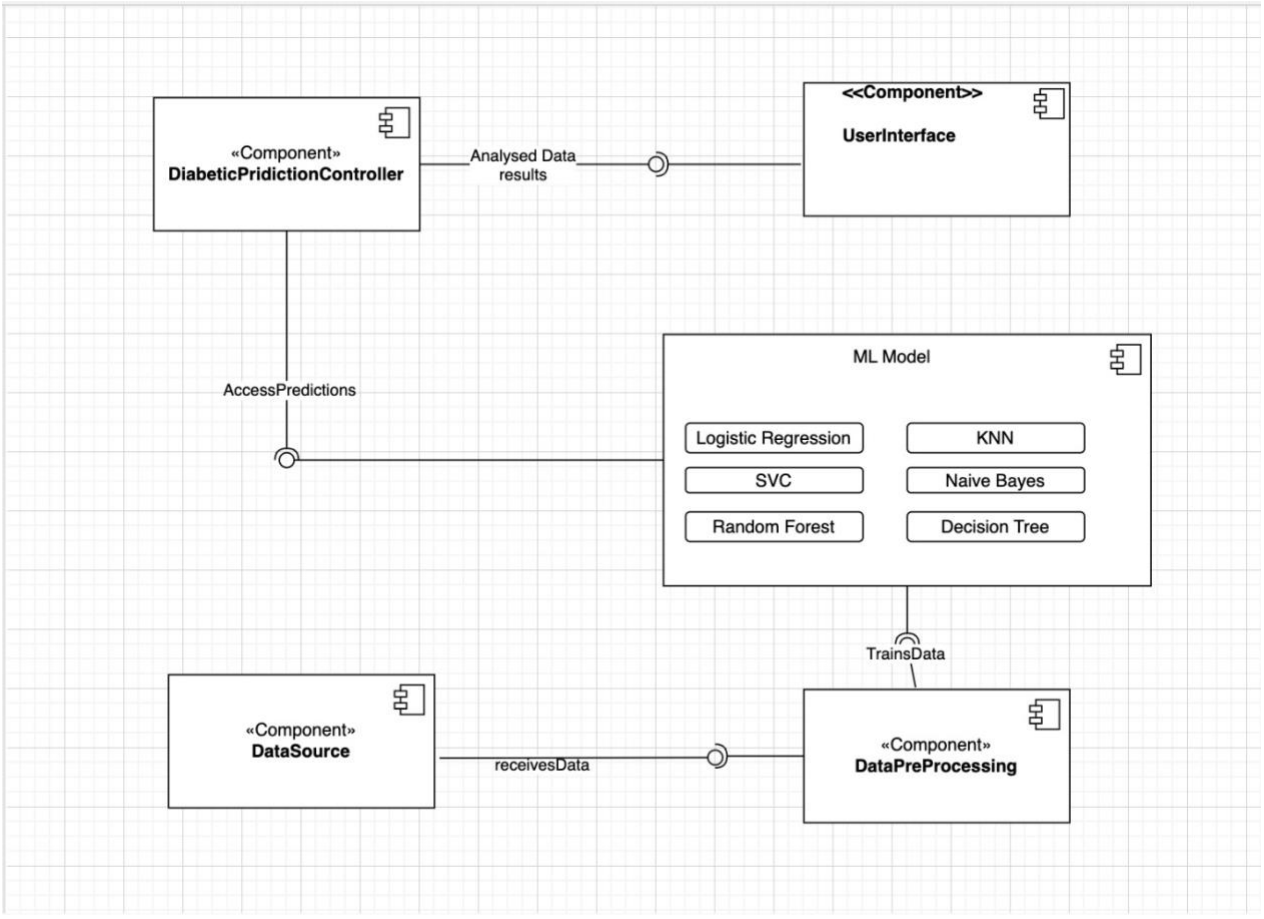
b) Educational content is added for the users to know about the relevant information.

c) Activity Tracker module and Personalized recommendation system are added to the system which tracks and gives tailored suggestions to user based on his/her activity.

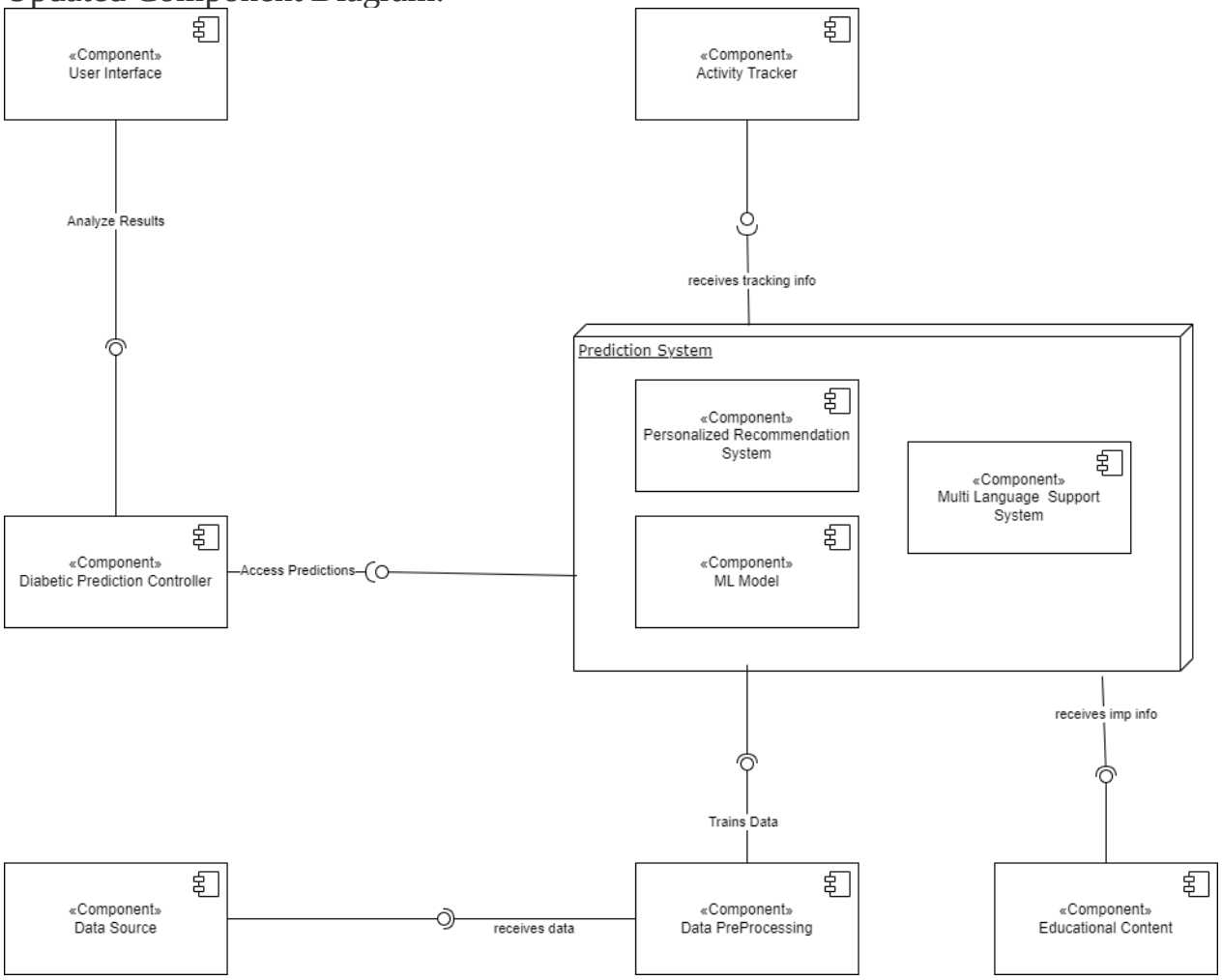
d) Multi Language Support is embedded into the system to be able to communicate in different languages.

2.3. Component Diagram

Component diagram from phase 1:



Updated Component Diagram:



Updates added are:

a) Educational content Component interacts with main model for the users to suggest about the relevant information.

b) Activity Tracker component and Personalized recommendation system components are added to the existing components for tracking and giving tailored suggestions to users based on their activity.

c) Multi Language Support component is integrated to the existing components making the system available in different languages.

Task 3: Reflection

The main lessons learned from this project are:

1. **Importance of Data:**
   Extensive, high-quality data is essential for precise risk assessment and tailored suggestions. Strong security and privacy protocols are essential for safeguarding user data. To increase model accuracy and adjust to changing patterns, ongoing data collection and analysis are necessary.

2. **Understanding User Needs:**
   To prevent misunderstandings, users need concise explanations, context, risk scores, and recommendations. Resources for education should be easily available and customized to meet the needs of a wide range of learners. System design must consider the possibility of bias and discrimination as well as user privacy concerns.

3. **Technology Considerations:**
   Selecting the right technology and architecture patterns can have a big impact on security, scalability, and performance. Event-driven and microservices architectures can improve system flexibility and future-proof it for changing requirements. Standardization and safe data sharing mechanisms are necessary for integration with wearable technology and health data systems.

4. **Ethical and Social Implications:**
   Care should be used when designing gamification components to encourage good habits rather than addictive or unhealthy competition. Users should not be discouraged or replaced by the system when seeking professional medical advice and interventions. Accessibility and digital literacy initiatives are crucial In order to guarantee equal access to healthcare technology and avoid escalating already-existing health inequities, accessibility and digital literacy programs are essential.

5. **Continuous Improvement:**
   To spot any problems and areas for development, system performance and user behavior must be tracked and recorded. To guarantee the efficacy and user acceptability of new features and gamification components, A/B

testing enables controlled evaluation. Interacting with stakeholders and medical experts yields insightful input for developing and executing systems.

To conclude, developing a diabetes prediction system highlights the necessity of a well-rounded strategy that incorporates ethical issues, user-centered design, and continuous cooperation with medical experts with cutting-edge technology.