State of the art

Mobile health checking is a huge market nowadays, approaching a number of $300 billions. Used to help with diagnosis, information about different medication or with measurements of lab results, these applications can be split in two main categories: intended for patients or intended for doctors and other medical staff. Because our project is part of the latter category we will study the market to see what are the capabilities of the applications available currently.

Epocrates is one such application, used by professional medical staff to quickly get information about medicaments and patient measurements, while also having a feature that allows diagnosis for some given symptoms, yet, some of that functionality is really expensive, worse is that there is no restriction when it comes to users, anyone can download and use the application, which may lead to problems. UpToDate is a product that solves this issue. While it has free content for normal users, UpToDate restricts it's more relevant functionalities, only certified practitioners can access them.

Our project implies two main phases. The collection of existing medical records and the diagnosis of new incomplete records. For the first phase, a doctor would complete an in depth form with relevant data of past cases. The fields of the form are based on the existing categories of real medical records, while loose description will be written in a separate field which will undergo further processing. This processing consists of applying different machine learning algorithms in order to separate a large body of text in a few key attributes. This phase can be evaluated by the numbers of records that we have in storage and by the fidelity of the correlations between our data and the real data added by the doctors.

Only after we have a respectable amount of records in our database we should be able to use the second phase functionality effectively. This phase implies that the user fills up another incomplete form with measurements and symptoms. The application will then use a clustering algorithm to categorise the input with some already existing records, thus giving an exact diagnosis. There are a multitude of highly efficient clustering algorithm including:

* K-means algorithm. One of the basic clustering algorithms, used to categorise n entities to k centers of clusters. In our case these centers would all be different diagnostics. It's main drawback is that it works with a constant number of clusters (k).
* ID3. A machine learning algorithm used to generate a decision tree from a dataset.
* Agglomerative Hierarchical Clustering. Is the most common type of hierarchical clustering used to group entities in clusters based on their similarity.

As a solution for our data storage we will proceed with a MySQL based database, we chose to do this because of a few main reasons:

* **Transaction based queries**
* Given that MySQL transactions don’t commit in case an error would appear during it’s runtime, they provide a very secure way to store data, given that it can be considered a second layer of tries and catches that don’t affect the integrity of the data
* **High Availability**
* MySQL is designed to process millions of queries and thousands of transactions while ensuring unique memory caches, full-text indexes and optimum speed
* **Scalability**
* In order to have high availability for all our customers MySQL can be an easy to scale tool, that can be accessed using cloud methods, AWS offers some ways to scale the number of database masters and slaves that are working at a given time
* **Reliability**
* With the recent changes in how user data should be handled and the addition of the GDPR, MySQL ensures data security with its data protection features, data encryption prevents unauthorized viewing of data and SSH and SSL support ensure safer connections
* It also features a powerful mechanism that restricts server access to authorized users and has the ability to block users even at the man-machine level
* the data backup feature facilitates point-in-time recovery
* **Quick-Start**
* You can go from software download to complete installation in just 15 minutes
* very quick, regardless of the underlying platform
* **Relational engines**
* The InnoDB engine constitutes a very powerful tool, given that we want to check the identities of the persons adding our data, assign regions for the inserted observations; foreign key constraints to maintain data integrity; fine grained locking-mechanism

As a documentation step, we've read the following related articles:

* Diagnosis of Clostridium difficile-associated disease: examination of multiple algorithms using toxin EIA, glutamate dehydrogenase EIA and loop-mediated isothermal amplification.  
  ([https://search.proquest.com/openview/cf78fe8565023013b0150db59e4b8821/1?pq-origsite=gscholar HYPERLINK "https://search.proquest.com/openview/cf78fe8565023013b0150db59e4b8821/1?pq-origsite=gscholar&cbl=4969"& HYPERLINK "https://search.proquest.com/openview/cf78fe8565023013b0150db59e4b8821/1?pq-origsite=gscholar&cbl=4969"cbl=4969](https://search.proquest.com/openview/cf78fe8565023013b0150db59e4b8821/1?pq-origsite=gscholar&cbl=4969))
* Artificial Intelligence in Personalized Medicine Application of AI Algorithms in Solving Personalized Medicine Problems.

(<https://www.researchgate.net/profile/Jamilu_Awwalu/publication/282624363_Artificial_Intelligence_in_Personalized_Medicine_Application_of_AI_Algorithms_in_Solving_Personalized_Medicine_Problems/links/5958ac5b458515ea4c62af76/Artificial-Intelligence-in-Personalized-Medicine-Application-of-AI-Algorithms-in-Solving-Personalized-Medicine-Problems.pdf>)

* Diagnosis of thyroid disease using artificial neural network methods.

(<https://ieeexplore.ieee.org/abstract/document/1199031>)

* Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices.

(<https://www.nature.com/articles/s41746-018-0040-6/>)

* Proposed diagnostic criteria and nosology of acute transverse myelitis.

(<https://miami.pure.elsevier.com/en/publications/proposed-diagnostic-criteria-and-nosology-of-acute-transverse-mye>)

* Artificial neural networks for diagnosis of hepatitis disease.

(<https://ieeexplore.ieee.org/abstract/document/1223422>)

* AI (artificial intelligence) in histopathology--from image analysis to automated diagnosis.

(<https://journals.viamedica.pl/folia_histochemica_cytobiologica/article/view/4316>)

* How artificial intelligence can help detect rare diseases.

(<https://www.sciencedaily.com/releases/2019/06/190606133805.htm>)