Name of the candidate:	**Munazzah Tasleem**
Name of the supervisor:	**Dr. Romana Ishrat**
Name of the co-supervisor:	**Dr. Md. Imtaiyaz Hassan**
Centre:	**Centre for Interdisciplinary Research in Basic Sciences**
Title:	**Human Disease Database: A new integrating data-model and tool for processing disease information**

Abstract

A database plays pivotal role in managing and distributing information, it is formed by organized collection and refinement of useful information from various resources to a single platform. We have developed a database, named Human Disease Insight (HDI), which will be helpful to researchers and clinicians. It will also provide basic information to non-professionals for creating awareness, thereby reducing the chances of suffering due to ignorance. Here, our mission is to provide a comprehensive human disease database containing most of the useful information with extensive cross-referencing. This database is a knowledge management system that acts as a central hub to access information about human diseases, drugs and genes involved in various diseases. In addition, the HDI contains well classified bioinformatics tool with description. We have developed a tool ProAnalysis by integrating nine essential bioinformatics tools for quick retrieval of results for the gene/protein sequences involved in causing human diseases, these results can be used further to enhance diagnosis and treatment of disease.

We collected human disease information through various online resources and medical text-books. The human diseases are classified into 12 broad categories according to ICD-10. Each category is precisely infiltrated with number of diseases carrying optimum information about each disease that includes its: synonym, general description, pathogen, gene, clinical feature, pathway, investigation, treatment, drug, prevention, prevalence, risk factor and references. At present we have 625 human diseases in HDI. A drop down list is provided on each page of HDI for accessing diseases. Navigation on disease will give list of disease categories, navigating on a particular category will list the disease present in it, clicking on a particular disease will render the stored information.

Medical practitioners require the drug information also, for which we have collected information about drugs used for the treatment of human disease. The drugs have been classified according to the NLEM in India. We have incorporated the description about the drug, name of the disease/s in which it is used and links to external database (DrugBank) for obtaining detailed information about the drug. Currently the HDI holds information about 320 drugs. A drop down list for drug is allocated at each page of HDI. Navigation on drug will list the drug categories, navigation on a particular category will yield drugs present in it. On clicking the drug of interest, user will receive its description, the name of the disease/s in which it is used and links to DrugBank.

We have searched the genes that are involved in causing human diseases and have listed them in HDI. We have linked each gene to the diseases in HDI and to external databases such as NCBI and UniProt for accessing maximum information about them. Gene list is provided on each page, navigation on it will enlist the 1440 genes stored in HDI. Clicking on a particular gene will give the name of the disease/s in which the gene is involved and links to external databases.

We have collected information about bioinformatics tools and have grouped them in 4 major groups, each group has been divided in further sub-groups, and each sub-group has number of tool in it. At present we have briefly described and linked 84 tools to their respective homepage. We have provided brief description of the tool and link to its main page. Placing bioinformatics tools in groups will render researchers for systematic execution of the gene/protein analysis.

The emergence of computational tools has an impact on the ameliorating and advancing means of various discovering methodologies for disease diagnosis and treatment. Hence, we have integrated nine essential bioinformatics tools and databases in HDI that will lay the basis for better diagnosis of the disease and locating the drug target for improving the treatment. The developed tool is named "ProAnalysis". User can select the gene of interest from the drop down list at HDI and they can also submit the UniProt ID to get the results from the integrated tools in just one click. This procedure allows end user to gain expert-quality results without having to understand the underlying decision-making methods being applied.

HDI is designed using HTML, CSS, PHP and Javascript. The relational database is created at the back end to store the curated information using MySql RDBMS. The front end of the HDI is user-friendly that provides navigations for retrieving information, for creating account, for searching the database, for FAQ and help pages, for Life & style page, for posting advertisements, for giving feedback and for liking the HDI to social networking sites. To avoid any breach in the database, we have incorporated security at the front end as well as at the backend. We have developed the administrative panel for updating the database that can be accessed only by the administrator.

HDI is a unique human disease database, although there are number of existing databases that provide information for a particular category of disease and databases that are useful only for clinicians or researchers. Multiple databases have been created to address problems related to specific disease categories, which highlights the importance of bringing together information on all human diseases from various categories under a common platform for analyses. This issue has been resolved by the creation of HDI, which contains relevant information about human diseases from various categories, along with descriptions and cross-references of the genes involved and the drugs used to treat them. This database contains a well-classified list of computational tools, along with descriptions of these tools and cross-linking to their sources. The distinct characteristic of HDI is the inclusion of information on genes, drugs and tools to aid in the exploration of human diseases. Additionally, word-suggesting search engine for searching diseases/genes and drugs is provided. HDI allows the downloading and uploading of relevant content. The data in HDI are updated regularly. The purpose of developing HDI was to provide a simple solution to allow physicians, researchers and non-professional to extract information on human diseases, genes, drugs, bioinformatics tools and perform quick analysis using a tool. The extensively cross-referenced, unified information provided, together with the facility of analyzing gene/protein using an integrated tool and downloading and uploading content in a user-friendly manner, confers unique characteristics upon HDI. Thus, the concept of HDI is different from other available disease databases. This database can be accessed free of charge to retrieve data on diseases, drugs and genes as well as tool-related information. Human Disease Insight database is available on the URL: http://www.humandiseaseinsight.com