Development of a basic database for quick screening ofanti-amyotrophic lateral sclerosis druglike compounds.

ABSTRACT

ALS is recognized as a multi-system disorder involving glial cells, immune system dysfunction, and disturbances in axonal transport, mitochondrial function and neurotrophic support. development of database is very essential for screening and identification of potential chemical compounds against ALS.The objective of this study was to design and develop a comprehensive database search engine specifically tailored for efficient screening of potential anti-ALS compounds with the scope to accelerate the identification of drug candidates with therapeutic efficacy against ALS. The methods used include curation potential drugs, batch pharmacokinetics prediction, database design, and domain hosting. The result was a fundamental database (OTPD4ALS) for swiftly screening potential anti-ALS drug compounds. The OTPD4ALS database is accessible at https://otpd4als.vercel.app.

Keywords: ALS, anti-ALS compounds, chemical database, MongoDB, JavaScript, TypeScript, React.js, Next.js, Node.js,pharmacokinetics.

UNDER PEER REVIEW

1.0. INTRODUCTION

Amyotrophic Lateral Sclerosis (ALS), also known as Lou Gehrig's disease, is a progressive and devastating neurodegenerative disorder that affects the motor neurons in the brain and spinal cord [1,2]. The clinical presentation of ALS is characterized by a combination of upper motor neuron (UMN) and lower motor neuron (LMN) signs. UMN degeneration results in

spasticity and hyperreflexia, while LMN degeneration leads to muscle weakness, atrophy, and fasciculations. The involvement of both UMN and LMN pathways contributes to the wide range of symptoms experienced by ALS patients, including difficulty in walking, speech difficulties, swallowing problems, and eventually respiratory impairment [3,4,5].

ALS is recognized as a multi-system disorder involving glial cells, immune system dysfunction, and disturbances in axonal transport, mitochondrial function and neurotrophic support [6]. Approximately 90-95% of ALS cases are sporadic, meaning they occur without a clear family history, while 5-10% are familial, resulting from genetic mutations passed down through generations [5,7]. Genetic studies have identified several genes associated with familial ALS, including SOD1, C9orf72, TARDBP, and FUS, and these mutations disrupt cellular processes, leading to motor neuron degeneration and the development of the disease [5,8].

Current treatment approaches for ALS mainly focus on symptom management and slowing disease progression. Riluzole and Edaravone are the only FDA-approved drugs for ALS, but their effects are modest but do not halt disease progression completely [5,7]. Despite significant efforts in ALS research, effective treatments for the disease remain limited, necessitating innovative approaches to accelerate the discovery of novel therapeutics.Collaboration and data sharing within the scientific community are essential to overcoming the challenges of ALS drug discovery. Computational drug discovery has emerged as a promising alternative to expedite the identification of potential drug candidates.Cheminformatics involves the study of application of databases in execution of chemicalknowledge[9]. The drug database offers a wealth of information on potential drug targets and candidate compounds, enabling efficient screening and prioritization of molecules for further investigation.Thus, development of basic database is very essential for screening

UNDER PEER REVIEW

and identification of potential chemical compounds against ALS.The objective of this study was to design and develop a comprehensive database search engine specifically tailored for efficient screening of potential anti-ALS compounds with the scope to accelerate the identification of drug candidates with therapeutic efficacy against ALS.

2.0 MATERIALS AND METHODS

2.1 ASSEMBLING OF ANTI-ALS DRUGS

To identify potential drugs for ALS and understand their mechanisms of action, a search was conducted on the Open Target Database (www.platform.opentargets.org) using the term "Amyotrophic Lateral Sclerosis." The database provided a list of 147 drugs that have been studied or proposed for ALS treatment, along with information about the mechanisms through which each drug acts. This valuable information was obtained in JSON format.To gain a deeper understanding of each identified drug, the names of these compounds were further searched on the PubChem database (https://pubchem.ncbi.nlm.nih.gov/).

2.2. *IN SILICO* PHARMAKINETICS PREDICTION

The SMILES of each of the drugs were assembled and used for *in silico* ADME (Absorption, Distribution, Metabolism, and Excretion) prediction, on the SwissADME webserver www.swissadme.ch[10], which was done at default parameters, and the results were saved in CSV format.

2.3 PREPARATION OF DATABASE FILE

To facilitate the creation of a database for the ADME prediction results, the original file in CSV (Comma-Separated Values) format was converted to JSON (JavaScript Object Notation)

UNDER PEER REVIEW

format. This conversion process was accomplished using the JSON Converter webserver [11].

2.4 DATABASE DESIGN

MongoDB's NoSQL attributes, including its flexible schema, scalability, real-time access, efficient querying, and compatibility with visualization tools, make it a well-suited database platform for storing and presenting ADME prediction information in the context of drug discovery. The decision to use MongoDB aligns with the need for seamless integration, scalability, and real-time accessibility in handling ADME data for informed decision-making.The database for storing the ADME prediction information was designed, constructed, and implemented using MongoDB, a popular NoSQL database management system known for its scalability and flexibility (https://www.mongodb.com/docs/). MongoDB was chosen as the database platform due to its ability to handle large volumes of data and its schema-free structure, which allows for easy storage and retrieval of JSON data.

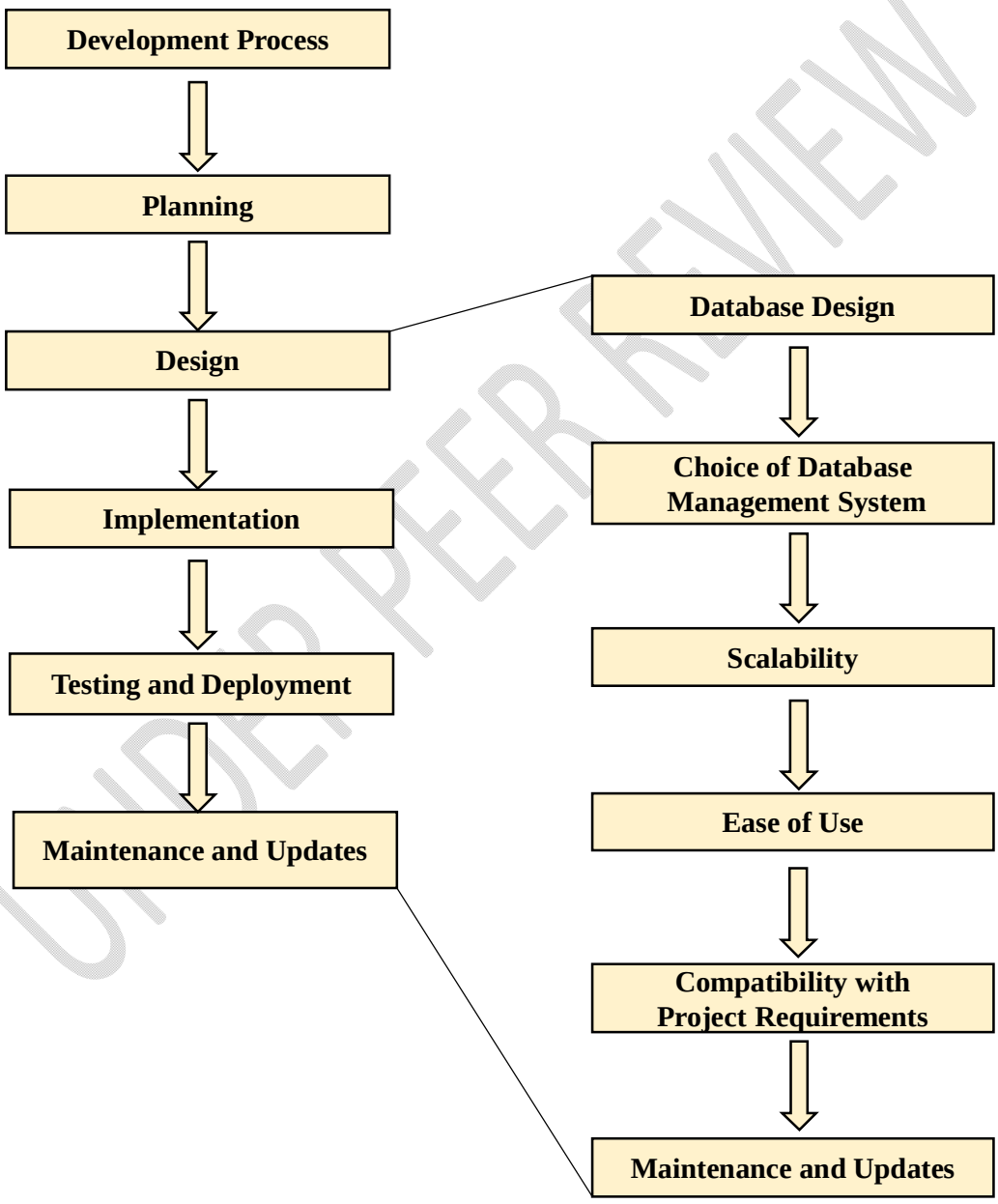
2.5 DOMAIN AND HOSTING

The domain and hosting of both the frontend and backend components of a website are crucial elements that determine how users can access and interact with the website's content and functionality. The fully completed database was used in the backend service which is hosted on an online server (railway.app) which is used to interact with the frontend which serves as the user interface for easy access which was also hosted on a free online hosting server (www.vercel.com).

2.6 USER FLOW AND USER EXPERIENCE(UX)

UNDER PEER REVIEW

User flow and user experience (UX) are critical aspects of web and application design that play a pivotal role in shaping how users interact with and perceive a digital product. These concepts encompass the journey users take through a platform and the overall quality of their interactions.The flowchart of the development of the database website is presented in Figure 1.



UNDER PEER REVIEW

Figure 1: Developmental flow of the database website

3.0 RESULTS

The OTPD4ALS database was created and it is accessible at https://otpd4als.vercel.app.

3.1 The @tanstack/react-table library

The @tanstack/react-table library was integrated during the implementation phase. This library empowered the creation of feature-rich tables in React.js applications. Equipped with sorting, filtering, and pagination functions, it proved ideal for showcasing extensive genomic data.The @tanstack/react-table library facilitated the creation of purpose-built tables for gene expression and genomic variations, allowing users to effortlessly organize, sort, and filter data. This enhanced the ability to discern complex patterns in the dataset.The synthesis of the @tanstack/react-table library with the web-based platform produced interactive avenues for exploring genomic insights. The tables were populated with attributes such as molecule, canonicalSmiles, esolClass, bbbPermeant, giAbsorption, cyp2d6Inhibitor, andlipinski.

3.2 Description of attributes

Molecule: A molecule is the smallest unit of a chemical compound that retains the chemical properties of that compound. It consists of atoms bonded together in specific arrangements. Molecules are the building blocks of matter and play a vital role in various biological and chemical processes.

Canonical SMILES: The Simplified Molecular Input Line Entry System (SMILES) is a notation used to represent molecular structures using ASCII characters. The Canonical SMILES representation provides a standardized way to encode molecular structures, making it easier to search and compare compounds in databases. It's especially useful in computational chemistry and cheminformatics.

UNDER PEER REVIEW

ESOL Class: ESOL (Estimated Solubility Class) is a classification system that predicts the water solubility of chemical compounds. It categorizes compounds into different classes based on their predicted solubility behavior. This prediction helps in drug formulation, as a compound's solubility can influence its bioavailability and efficacy.

BBB Permeant: BBB (Blood-Brain Barrier) Permeant refers to a compound's ability to pass through the blood-brain barrier, a selectively permeable membrane that separates the blood from the brain's extracellular fluid. A compound that is BBB permeant can potentially affect the central nervous system and is relevant for drug development targeting neurological disorders.

GI Absorption: Gastrointestinal (GI) absorption refers to how well a drug or compound is absorbed into the bloodstream after oral administration. It's a crucial factor in determining the bioavailability of orally administered drugs. Understanding GI absorption helps predict a compound's effectiveness and required dosage.

CYP2D6 Inhibitor: CYP2D6 is an enzyme responsible for metabolizing many drugs. A CYP2D6 inhibitor is a compound that interferes with the activity of this enzyme, affecting the metabolism of drugs that rely on CYP2D6 for processing. Inhibition can lead to altered drug levels in the body and potential drug-drug interactions.

Lipinski Rule of Five: The Lipinski Rule of Five, developed by Christopher Lipinski, is a set of criteria used to evaluate the drug-likeness of chemical compounds. It includes factors such as molecular weight, lipophilicity, hydrogen bond donors and acceptors, and more. Compounds that adhere to these rules are more likely to have favorable pharmacokinetic properties and are considered potential drug candidates.

UNDER PEER REVIEW

Each of these attributes plays a unique role in characterizing and evaluating potential drug-like compounds for the treatment of ALS. Their combined insights contribute to the systematic screening and identification of compounds with desirable properties for further investigation and development.The results of query are presented in tabular or graphical visualization, with subsequent sections elaborating on the significance of each attribute (**Figure 2**).

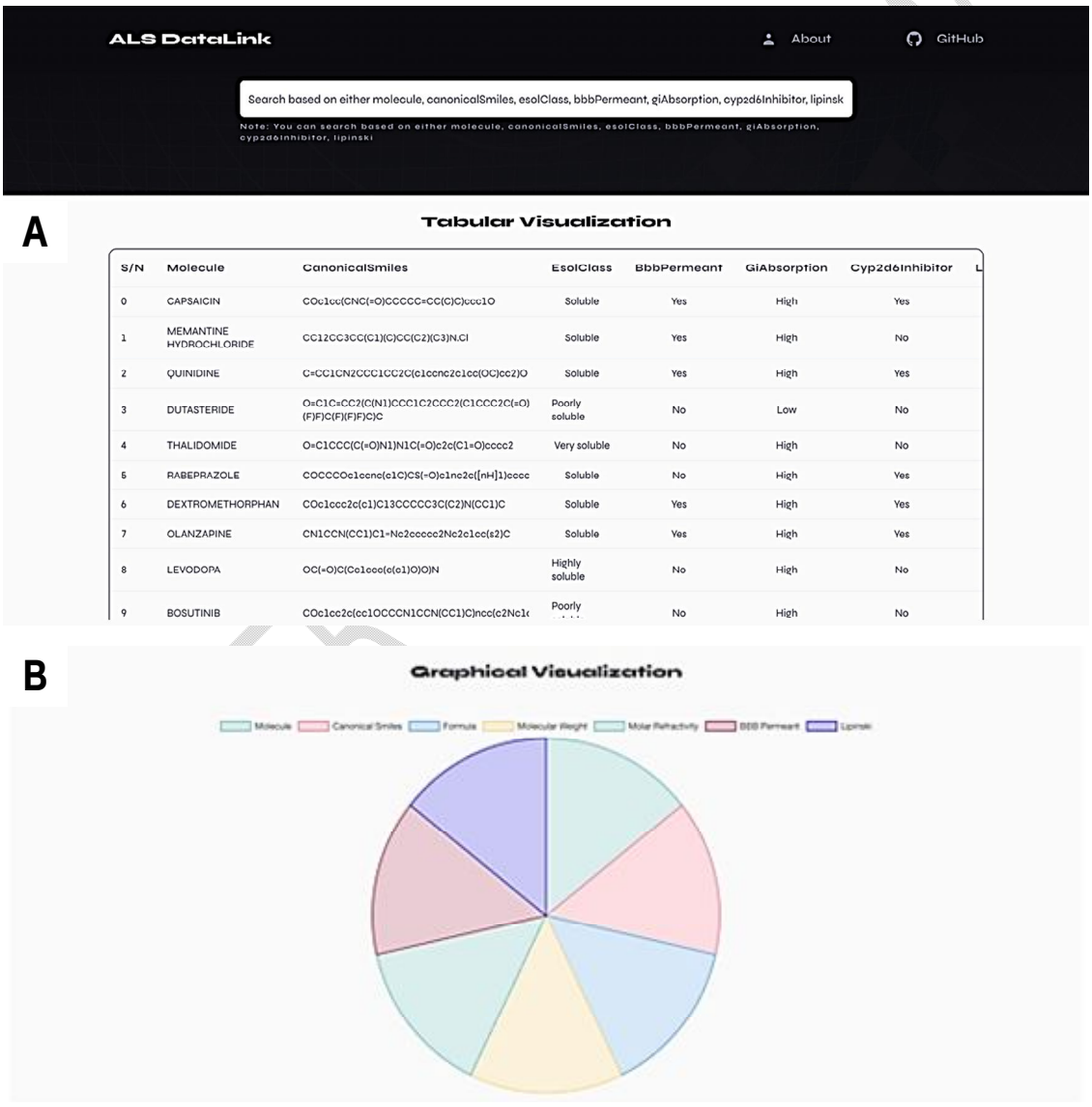
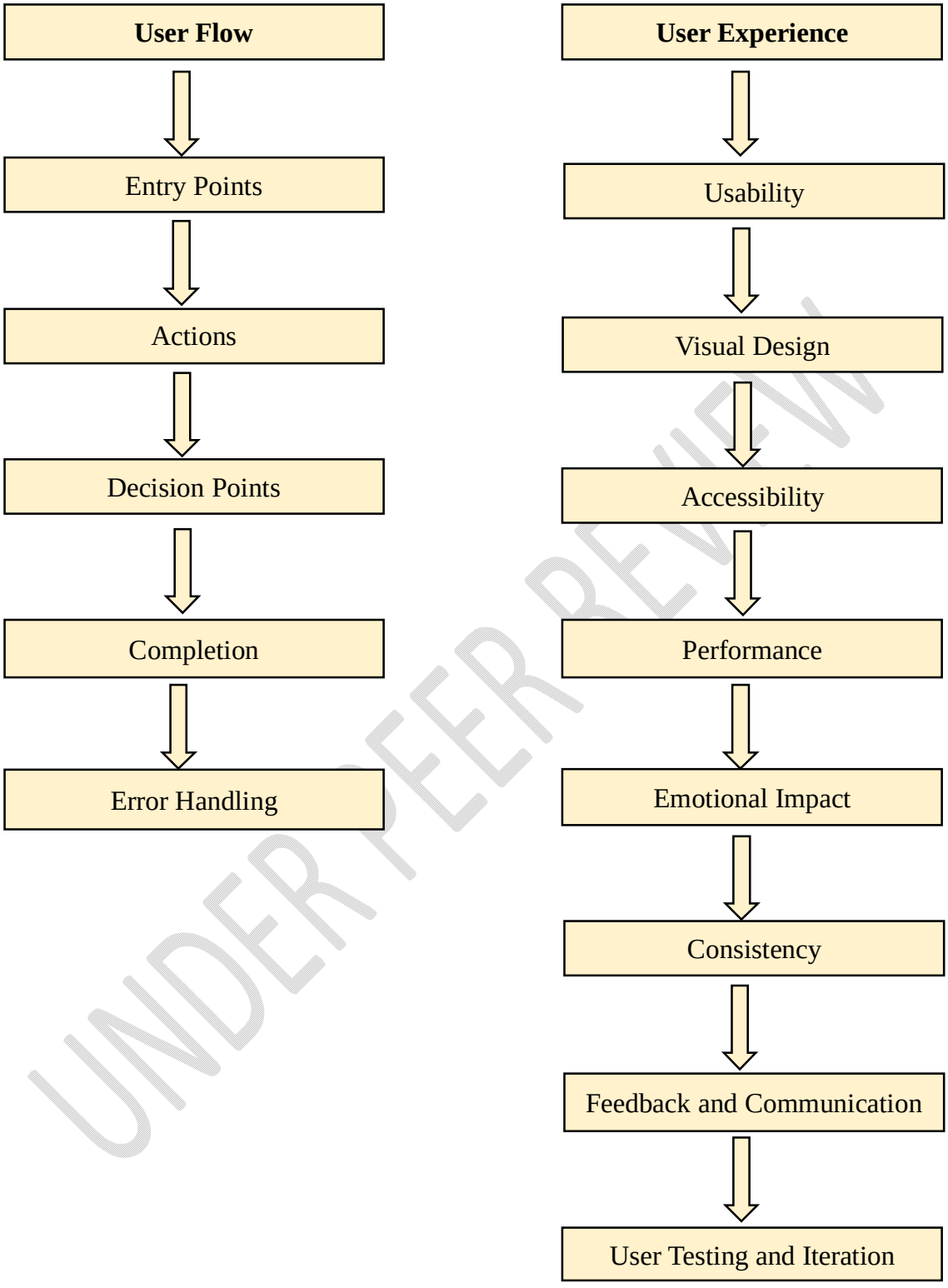


Figure 2: Tabular and graphical visualization of data on the database website

3.3 User flowand user experienceintegration

User flow and user experience (UX) are interconnected (Figure 3). A well-designed user flow contributes to a positive user experience by guiding users through the platform efficiently. On the other hand, a strong UX design enhances user flow by providing intuitive interactions and reducing friction during the user journey.User flow outlines the sequence of steps users take to achieve their goals within a platform, while user experience encompasses the overall quality of these interactions. A seamless integration of user flow and UX design results in a user-friendly, accessible, and engaging digital product that meets users' needs and expectations.

UNDER PEER REVIEW



UNDER PEER REVIEW

Figure 3: User Flowand User Experience of the Website

4.0 DISCUSSION

This study has delved into the practical demonstration of a fundamental component crucial to the endeavour of identifying anti-ALS druglike compounds by development of a basic yet efficient database system. The purpose of this database is to provide a platform that enables researchers and scientists to rapidly screen and evaluate potential compounds for their suitability as candidates in combating ALS.

The development of a drug database specifically tailored for ALS, such as the "OTPD4ALS," addresses the multifaceted nature of the disease. ALS is characterized by the involvement of various biological systems [12]. A database dedicated to ALS must therefore integrate diverse data to facilitate the identification of compounds that can effectively target these different pathological aspects.Comprehensive drug databases like DrugBank or ChEMBL have been instrumental in drug discovery by providing access to vast amounts of data, including drug interactions, pharmacodynamics, and pharmacokinetics [13,14]. These databases allow researchers to screen for compounds with specific properties or actions against known targets, thereby streamlining the drug discovery process. The creation of "OTPD4ALS" similarly aims to provide a resource tailored to ALS research, enabling efficient screening of potential anti-ALS compounds.

A critical component of any drug database is the curation of compounds. This involves collecting and verifying data from various sources to ensure that the database contains accurate and up-to-date information. For ALS, curating compounds that affect the disease's pathophysiology such as those targeting neuroinflammation, oxidative stress, or mitochondrial dysfunction, is particularly important. Additionally, predictive tools integrated into databases enhance their utility by allowing for in silico analysis of drug-likeness, pharmacokinetics, and toxicity. These tools can predict how a compound might behave in a

UNDER PEER REVIEW

biological system, which is crucial for early-stage drug discovery. The integration of pharmacokinetics parameters in "OTPD4ALS" exemplifies this approach, enabling researchers to quickly filter out compounds that are unlikely to be viable as drugs due to poor bioavailability or non-BBB permeability [2,7].

The accessibility of these databases via web platforms is another significant advantage. By hosting "OTPD4ALS" online, the database becomes a global resource that can be accessed by researchers from around the world, fostering collaboration and accelerating research. This is particularly important in rare or complex diseases like ALS, where pooling resources and data can lead to more significant breakthroughs [15].Despite their benefits, drug databases also face several challenges. These include the need for continuous updatesand the difficulty of integrating diverse types of biological data into a cohesive platform. The database may be limited to currently known and published compounds, potentially overlooking novel or emerging compounds that have not yet been documented or included.For databases focused on diseases like ALS, where the disease mechanisms are still not fully understood, these challenges are particularly pronounced. Future developments should focus on improving data integration, enhancing predictive models, and ensuring that the databases are regularly updated with new research findings.

5.0 CONCLUSION

In conclusion, the development of a basic database for quick screening of anti-ALS drug-like compounds represents a significant step towards advancing ALS research and potential therapeutic interventions. The database serves as a valuable resource, integrating genetic information, chemical properties, and drug-target interactions. This approach enhances the efficiency of identifying potential drug candidates and accelerates the drug discovery process for ALS.The accessible nature of the OTPD4ALS website at

UNDER PEER REVIEW

https://otpd4als.vercel.appromises to accelerate the discovery and development of effective treatments for ALS, ultimately advancing efforts in combating this challenging neurological condition.While challenges and limitations exist, a concerted effort to improve and refine the database will undoubtedly contribute to advancing our understanding of ALS and bringing us closer to transformative therapies for people affected by this debilitating disorder.

UNDER PEER REVIEW

REFERENCES

1. Brown, R. H., & Al-Chalabi, A. (2017). Amyotrophic lateral sclerosis. New England Journal of Medicine, 377(2), 162-172.

2. Fatoki TH, Saliu IO, Balogun TC, et al. (2024). In silico investigation of lipid-based compounds implicated in amyotrophic lateral sclerosis. Brain & Heart doi: 10.36922/bh.2976

3. Swash, M., & Schwartz, M. S. (2000). The clinical spectrum of amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders, 1(5), 287-290.

4. Kiernan, M. C., Vucic, S., Cheah, B. C., Turner, M. R., Eisen, A., Hardiman, O., et al. (2011). Amyotrophic lateral sclerosis. The Lancet, 377(9769), 942-955.

5. Fatoki T.H., Chukwuejim S., Udenigwe C.C., and Rotimi E. Aluko RE., (2023). In silico exploration of metabolically active peptides as potential therapeutic agents against amyotrophic lateral sclerosis. International Journal of Molecular Sciences, 24(6): 5828.

6. Al-Chalabi, A., van den Berg, L. H., &Veldink, J. (2017). Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. Nature Reviews Neurology, 13(2), 96-104.

7. Fatoki T., Chukwuejim S., Ibraheem O., Oke C., Ejimadu B., Olaoye I., Oyegbenro O., Salami T., Basorun R., Oluwadare O., Salawudeen Y., (2022). Harmine and 7,8-dihydroxyflavone synergistically suitable for amyotrophic lateral sclerosis management: An insilico study. Research Results in Pharmacology, 8(3):49-61.

8. Renton, A. E., Majounie, E., Waite, A., Simón-Sánchez, J., Rollinson, S., Gibbs, J. R., ... & Traynor, B. J. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron, 72(2), 257-268.

9. Wang, H. (2008). Design of a Structure Search Engine for Chemical Compound Database. PhD Dissertation, Georgia State University. doi: https://doi.org/10.57709/1059443

10 Daina, A., Michielin, O., & Zoete, V. (2017). SwissADME: A free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. Scientific Reports, 7, 42717.

11 JSON Converter webserver. Available at https://products.groupdocs.app/conversion/csv-to-json. Accessed on: May 30, 2024

12 Boillée, S., Vande Velde, C., & Cleveland, D. W. (2006). ALS: A Disease of Motor Neurons and Their Nonneuronal Neighbors. *Neuron*, 52(1), 39-59. doi:10.1016/j.neuron.2006.09.018.

13 Wishart, D. S., Knox, C., Guo, A. C., Cheng, D., Shrivastava, S., Tzur, D., ... &Orehov, G. (2008). DrugBank: a knowledgebase for drugs, drug actions, and drug targets. *Nucleic Acids Research*, 36(Database issue), D901-D906. doi:10.1093/nar/gkm958.

14 Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., ... &Overington, J. P. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Research*, 40(Database issue), D1100-D1107. doi:10.1093/nar/gkr777.

15 Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., Liu, Y., et al. (2014). DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Research*, 42(Database issue), D1091-D1097. doi:10.1093/nar/gkt1068.

UNDER PEER REVIEW