JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 09 Issue 01
2025

Research Article
https://doi.org/10.56979/901/2025

LncRNAs Disease: A Text Mining Approach to Find the Role of LncRNA in Aging

Hira Shakeel[1], Misbah Akram[2], Muhammad Usman Javeed[1], Muhammad Azhar[3*], Shafqat Maria Aslam[4], Saifullah[5] and Muhammad Tayyab Mumtaz[1]

[1]Department of Computer Science, COMSATS University of Islamabad, Sahiwal, 5700, Pakistan.
[2]Department of Software Engineering, Minhaj University Lahore, Lahore, 54000, Pakistan.
[3]Department of Applied Data Science, Hong Kong Shue Yan Unversity, SAR, China.
[4]School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi, 710062, China.
[5]Department of Zoology, Wildlife and Fisheries, University of Agriculture, Depalpur Campus, Pakistan.
[*]Corresponding Author: Muhammad Azhar. Email: azhar@hksyu.edu

Received: April 17, 2025 **Accepted:** May 31, 2025

Abstract: The number of studies on long non-coding RNAs (LncRNAs) has rapidly expanded in recent years. Since LncRNAs have been shown to have strong associations with a variety of illnesses, it is critical to comprehend their function or roles in disease diagnosis, prognosis, and therapy response assessment. Using keywords like "Long non-coding RNAs" and "LncRNAs," we searched through over 95,000 papers and carefully examined them to get information about the relationship between LncRNAs and diseases. Because several LncRNA identifiers (IDs) and reference genome versions exist, these data are highly synthesised and diverse. We normalized the data by making a significant effort to eliminate anomalies and repetition. In the end, we created the LncRNAs Disease database, which incorporates data from experiments supporting the association of LncRNAs with a variety of illnesses. A unique database resource, The LncRNAs Disease (LncRNA Disease v1.0) offers over 90000 hand selected connections of 57096 empirically supported LncRNAs implicated in 28 disorders. By offering comprehensive details on the LncRNA in each illness, together with experimental support, a concise description, sequence information, and location details, LncRNAs illness helps users. Given the function or roles of LncRNAs in a variety of disorders, it is anticipated that a vast amount of data would be generated soon. As a result, we provide a submission area where researchers or anyone may help us update our database on long non-coding RNA diseases.

Keywords: Long non-coding RNAs; LncRNAs; Textmining LncRNA; Aging LncRNA; Senescence LncRNAs

1. Introduction

In combination with the conventionally annotated protein-encoding genes (ncRNAs), genomes also produce hundreds of regulatory non-coding RNAs. ncRNAs play a part in the control of transcriptional and post-transcriptional processes as well as the activity of chromatin-modifying complexes. There are several forms of non-coding RNA, but in recent years, long non-coding RNA and microRNAs have drawn a lot of interest [1]. Genomes create tens of thousands of regulatory non-coding RNAs (ncRNAs). Examples of "housekeeping" ncRNA transcripts include long non-coding RNAs, microRNAs, piRNAs, natural antisense transcripts (NATs), and a number of other little-known ncRNAs that originate from the transcription of expression control sequence elements (ribosomal RNA, transfer RNA, small nuclear RNA, and small nucleolar RNA) [2]. By exhibiting dynamic spatial and temporal expression patterns under certain physiological conditions, regulatory noncoding RNAs (ncRNAs) aid in tissue patterning and the regulation of several cellular programs, such as cell proliferation, differentiation, migration, and death.

Several are poorly defined kinds of non-coding RNAs such as, Ribosomal non-coding RNA, transfer non-coding RNA, small nuclear non-coding RNA, and small nucleolus RNA, regulatory ncRNAs is added to expressed lower RNAs [3]. Small non-coding RNAs < 200 bps, e.g. miRNAs, siRNAs, and piRNAs, and LncRNAs, microRNAs are two types of regulatory ncRNAs that is long ncRNAs ≥ 200 bps [4]. In the last few decennium of RNA biology research, several LncRNAs such as Xist and H19 have been identified and are considered highlights in LncRNA biology. With the initiation of highly developed sequencing techniques and insights from huge association paying attention on the characterization of purposeful genomics fundamentals such as ENCODE (Encyclopedia of DNA Elements), more and more long ncRNAs have been recognized and await functional legalization increased. The regulation of chromatin composition, transcription, and post-transcriptional levels are concerned with LncRNAs and the understanding of eukaryotic genome structure, activity, and regulations have been shown to transfigure. Genomic complication have added another level of LncRNAs.

Scientists now know exponentially more about LncRNAs than we did ten years ago when they were discovered. Particularly common in the neurological system, LncRNAs are particularly cell and tissue specific. LncRNAs are impacted by intracellular and intercellular trafficking as well as post-transcriptional processing [5]. Utilizing their capacity to engage in conformational interactions and sequence-specific with a wide range of partners (DNA, RNA, and proteins), LncRNAs operate through a number of molecular mechanisms. Due to their tiny size, LncRNAs work in a modular fashion, integrating many macromolecules in the cell's three-dimensional environment. Thus, LncRNAs coordinate the execution of critical cellular programs including development, growth, cell identity, and stress response deployment as well as transcriptional, post-transcriptional, and epigenetic activities. According to new study, In mediating the brain's cellular diversity LncRNAs play a significant role as well as in developmental complexity and activity-dependent plasticity. Other studies have connected these characteristics to senescence and brain aging [6].

The LncRNAs that have been discovered Act via a remarkably varied spectrum of processes that have been examined transcriptional, post-transcriptional, and epigenetic regulation, as well as other molecular processes [7]. However, for the vast majority of LncRNAs, their biological functions are unknown. The complete significance has yet to be determined. Some LncRNA genes have been considered to be significantly conserved throughout evolution. Figure 1. Illustrate the intergenic and intronic LncRNA.

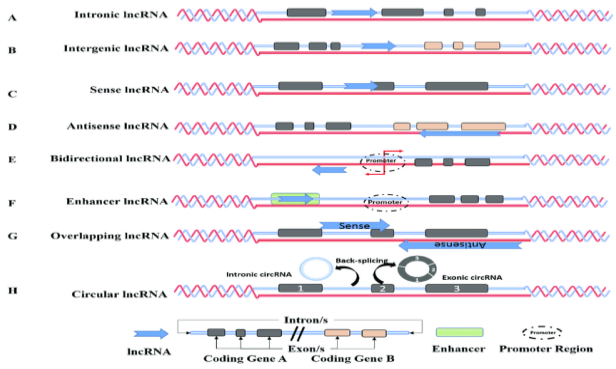


Figure 1. Intergenic and Intronic LncRNA

If required, please draw attention to contentious and opposing theories. Last but not least, briefly state the work's primary goal and emphasize its key findings. To the best of your ability, make the introduction understandable to scientists who are not in your scientific specialty. References should be inscribed with a number or numbers in square brackets, such as [1], [2,3], or [4–6], and numbered according to appearance. For further information about references, see the document's conclusion.

Their vital roles in cellular activities are being supported. Others In a lineage-specific manner, LncRNA genes have developed quickly manner, suggesting that they acted as foundations for evolutionary advances, including the origin of human brain Higher-order neurobehavioral functions, as well as structure. Long non-coding RNAs (LncRNAs) are emerging as important mediators in a variety of physiological and pathologic processes. Despite the fact that a large number of LncRNAs have been discovered, only a fraction has been functionally defined in ageing. We examined at genome-wide LncRNA expression

during cellular senescence using human fibroblast cells. Senescence is the term describing the restricted capacity of human fibroblasts to divide in culture. After that, they survive for a few weeks but do not multiply even when given space, food, and growth hormones. Telomere erosion (replicative senescence) and exposure to harmful circumstances are two closely related processes that can be cause of cellular senescence (premature senescence).

Senescent cells feature are an expanded shape, increased lysosomal-galactosidase activity, increased autophagy, and senescence-associated heterochromatic foci (SAHF). The senescence-associated secretory phenotype is also seen in senescent cells. Cellular senescence has been demonstrated to influence other diseases such as neurodegenerative, cardiovascular disease, and deteriorating immunological function, in addition to acting as a crucial anti-cancer strategy. Senescence regulation has been linked to a variety of proteins, including short non-coding RNAs, transcription factors from the p53 and Ets families, and post-transcriptional regulators AUF1 (AU-binding factor 1) and HuR (human antigen R). According to a recent study, long non-coding RNAs (LncRNAs) exhibit a unique pattern of expression during cellular senescence and are probably a crucial regulator of cellular senescence.

A diverse group of transcripts, LncRNAs may be found in a range of subcellular locations, sizes, and forms. LncRNAs can regulate gene expression via changing the makeup of transcription factors and chromatin reconfiguration, as well as by producing DNA-RNA triple helices. In furthermore, LncRNAs have a scaffolding or decoy function that affects gene transcription. Furthermore, LncRNAs influence gene expression post-transcriptional via regulating precursor mRNA splicing, mRNA stability, or translation. Cell proliferation, differentiation, chromosomal imprinting, and embryogenesis are all affected by LncRNA-mediated gene expression, On the other hand, LncRNA function is associated with metabolic disorders, cancer, neurodegeneration, and cardiovascular disease. Despite the fact that LncRNAs are increasingly being found in a wide range of physiological and pathologic processes, little is known about how they function as people age.

In the aged population, vascular ageing is a major source of morbidity and mortality. Endothelial cells (ECs) and vascular smooth muscle cells (VSMCs), which compose the intima and media layers of the artery wall, are intimately linked to the ageing process and vascular aging-related illnesses. Numerous research have demonstrated the pathophysiologic mechanism by which LncRNA contributes to vascular ageing, therefore antisense long non-coding RNA (AS-LncRNA) is currently receiving more attention in the pathogenesis of vascular ageing. Despite this, only a few researches have focused on the exact mechanism through which AS-LncRNA mediates vascular ageing.

Long non-coding RNAs (LncRNAs) are transcripts that have low potential to code for proteins yet responsible for a considerable portion of transcriptional output. In cellular homeostasis, they regulate gene expression at the epigenetic, transcriptional, and post-transcriptional mechanisms. However, LncRNA research is still in its infancy, and the great majority of LncRNA transcripts' functions are uncertain. It is commonly recognized that the precise control of gene expression is essential to the functioning of the human nervous system. Numerous studies have shown that LncRNAs significantly affect both the onset and course of neurodegenerative disorders as well as normal brain development. We evaluated recent studies on the role of LncRNAs in neurodegenerative illnesses, including glaucoma, multiple system atrophy (MSA), front temporal lobar degeneration (FTLD), amyotrophic lateral sclerosis (ALS), Huntington's disease (HD), Parkinson's disease (PD), Alzheimer's disease (AD), and multiple system atrophy (MSA).

Glaucoma, which is characterized by unexplained ganglion cell damage and death, is now classified as a chronic neurodegenerative diseases, thus we also talked about LncRNAs and glaucoma. We demonstrate the function of a few particular LncRNAs, which may provide fresh insights into the etiologic and path physiology of the neurodegenerative diseases highlighted above. Because they suppress insertion mutations caused by transposable elements, LncRNAs serve a critical role in protecting the genome, maintaining its complexity and integrity. Previously, the significance of LncRNAs was thought to be limited to gonad development, The brain, colon, heart, kidney, liver, lung, small intestine, spleen, stomach, ovary, and testis are among the organs with LncRNA expression patterns, according to current studies.

Therefore, LncRNAs play an important role in disease progression, diagnosis, and treatment response evaluation. The expression of LncRNAs is deregulated in a variety of diseases, according to genome-wide profiling studies. However, target-based mechanistic studies have demonstrated that LncRNAs have a regulatory function in a variety of diseases. Target genes are regulated by LncRNAs through a base paring

method. Emerging data suggests that changes in LncRNA expression and abnormalities in target gene regulation could be used as a diagnostic marker. In order to offer fundamental information on LncRNAs, a number of databases have been built recently, including LncRNABank and lncRBase, which provide extensive LncRNA sequence and location information for numerous species. The diversified story of LncRNA Quest, another database resource, focuses on pseudo genes and system information, including sequencing and location data.

However, a number of databases demonstrate the connection between illness and non-coding RNAs, including short and long non-coding RNAs. Among the databases are circRNAdisease, circ2disease, miR2disease, miRCancer, LncRNADisease, and Lnc2cancer. Nevertheless, no online database resource exists that offers details on the connection between LncRNAs and illness. Consequently, In addition to empirically verified lncR-NAs, we created a manually curated LncRNA and illness association database resource that includes disease association linkages from the literature.

2. Related Work

Bioinformatics has received accumulating attention throughout in the world because of that the biomedicine and technology prog ression have flourished. In bioinformatics, they are doing no longer encipher super molecule arrangement are frequently replicating as non-coding RNAs (ncRNAs) in the region of the human genome [8]. The kinds of non-coding RNAs are primarily supported the period of transcripts of non-coding RNAs, small non-coding RNA, and long non-coding RNAs. LncRNAs are furthermore 200 nucleotides in length and create up the majority of non-coding RNAs and between the long non-coding RNA and small non-coding RNA have a clear difference.

In recently past few years, LncRNA have received a great deal of attention from bioinformatics and researchers. Highly exaggerated shreds of proof suggest that LncRNAs ordinarily performed in cancer or neoplasma suppressor functions in disease of human cancers [9]. Including prostate cancer [10], heap to cellular carcinoma (HCC), breast cancer, colon cancer, aging, liver cancer, lung cancer [11], bladder cancer [12], epigenetic [13] and others diseases. Even now, countless times new LncRNAs are revealed in every year. By using the biological experimental methods, the long non-coding RNAs increased their amount and makes it more complicated to identify the association of LncRNA diseases. Due to the time of the usage of biological experiments is to perceive the affiliation of LncRNA diseases ends in a bottleneck and the experiment values are concerned within sides.

As a result, the screening area for biological experiments can effectively reduce their computational techniques to predicting the probable lncRNA disease associations,the time and cost of biological experiments are moderated. In addition, predictive calculations can help you to discover the cause and mechanism of the disease as soon as possible. This mechanism is extremely necessary for diseases to diagnosis, drug prognosis, and target discovery [16]. The transcription factors (TFs) could hold collectively with LncRNAs promoters and at the DNA level be controlled by TFs. Pubmed data is available in different sites just like as, The National Center for Biotechnology Information, available information about non-protein-coding RNA genes, and provides the data on genes–PubMed relationships. Secondly, authors manually vetted the data and extricate the LncRNAs disease in the form of pairings. All LncRNAs disease relationships are cross-checked by different bio informatitions and researchers. The source articles in the Pub Med databases were linked. The sequencing and species data annotations have also added by the authors. The names of LncRNAs, aging, and diseases were also homogenized. The authors are curated 166 illnesses in all, added the different types of cancer, aging, epigenetic, and cardiovascular disease, and neurodegenerative disease are ranking first, second, and third, respectively.

Databases have been compiled that relates with assorted LncRNA disease. In 2013 LncRNA Disease [14], association database is founded and in that field it was the first database that have been created. In 2015, Lnc2Cancer was founded. This dataset first and foremost contains the associations of data in between epigenetic, cancer and LncRNA. It is associated to LncRNAs Diseases; the entry of Lnc2 Cancer is broader and more complicated. NONCODE is a inclusive knowledge of the database restrain virtually all the ncRNAs, and LNCipedia is a wide-ranging human LncRNA database. In adding up, the protein-coding function make available different tools for predicting.

Predict the association of a new form of human LncRNA disease. Automatically predicting an IncRNA disease association was an initial study. Chen Xiang has since created up some enhancements that supported the LRL SLDA model [15]. Researchers proposed them a global (RWRlncD) network-based computing framework that predicting the potential of LncRNA disease associations by accomplishing a restart random walk (RRW) method on LncRNA functional resemblanced networks [3]. To predicting the potential IncRNA disease associations thus the hypergeometric distribution model (HGLDA) is newly developed model by researchers. Zhou et al. propose an RWRHLD method to assimilate the miRNAs-related LncRNA, siRNA therapeutics, and LncRNAs crosstalk networks, disease-like networks, and well-known LncRNA disease-related networks into newly developed networks predicting the potential of IncRNA disease associations based on integrated networks [16, 17].

Investigate the disease apparatus and functions of LncRNAs are those methods available for relating the data. However, some uncertain blocks are still having present in models. They are having complexities and it is very complex to calculate and neglects to select parameters. Therefore, to predict the association of IncRNAs disease much research has stay behind to be done [18].

The IncRNA disease consequences have been shown solid results because of that the endure methods of prediction are available, but having a certain limitation, there is still some rooms for improvement. Those technologies are supported a new automated methods by Random Forest (RF) or Incremental Principal Component Analysis (IPCA) and techniques are build up for predicting the long non-coding RNA disease associations [19, 20]. Those are named this IPCARF. First, the semantic similarity of the disease has been included, in the long non-coding RNAs added well-designed resemblance, and the similarities of the Gaussian interaction spectrum to obtain the characteristics vector of the long non-coding RNA diseases pairs. The feature and dimensions of the dataset by using the IPCA method to effectively reducing and achieve the optimal features for sub spacing from the original features set [21]. In the end, they have trained the RF model to predict the potential IncRNA disease associations.

Deregulation of LncRNAs has lately been linked to a va- riety of human diseases, most notably diseases, neurological disorders, and cardiovascular diseases, thanks to advances in global transcript me profiling techniques. Some LncRNAs are increasingly being implicated in cancer progression, including proliferation, invasion, and metastasis. As a result, LncRNAs may be considered a viable cancer prognostic marker. Accord- ing to a sketch of the human genome project (HGP), there are only about 20,000 protein-coding genes in the human genome, accounting for less than 2. LncRNAs are now characterized as a huge and heterogeneous class of transcribed RNA molecules that lack a major open reading frame and are longer than 200 nucleotides in length. The preponderance of LncRNAs are RNA polymerase II transcripts with a poly-A tail and cap, similar to protein-coding RNAs. Interestingly, most LncRNAs are primarily found in the nucleus of the cell, and they have lower evolutionary conservation or levels of expression than mRNAs. Many studies have already shown that LncRNAs can control gene expression levels, post-transcriptional changes, bind to transcription factors or miRNAs, and function as a modulator in a variety of biological processes. LncRNAs are predicted to increase our understanding of disease progression and reveal potential biomarkers for diagnosis and prognosis as a result of their unique RNA characteristics, more tissue-specific expression fashion, and more stable structure.

Previous research has demonstrated that LncRNAs can improve mRNA stability and consequently regulate mRNA expression [22]. As a result, one reasonable interpretation is that the competing partners, such as sponge LncRNAs, trigger the expression levels of these discharged mRNAs. An extreme form of somatic mosaicism develops from the accumulation of genetic and epigenetic changes, which can lead to cancer.

Although the history of cancer is always evolving, it is still a complicated disease that requires a high level of expertise. Up until recently, much of the causal evidence for the onset and spread of cancer has been connected to regions that code for proteins. Ultra-conserved non-coding sequences, on the other hand, are frequently shown to be deregulated in cancer. 9 Single nucleotide polymorphisms (SNPs), for example, are among the high-risk modifications linked to the development of cancer; interestingly, 85 percent of SNPs are designated in noncoding areas and linked to disease development. 10 These abnormalities have an effect on LncRNA, which have changed expression and functions, resulting in deregulation of their targets.

Long non-coding RNAs (LncRNAs) are RNAs with transcripts longer than 200nt but no specific open reading frames, making them unable to encode proteins. LncRNAs were formerly considered to be transcript me to noise or garbage sequences produced by RNA polymerase II (Pol II), and researchers ignored them. In reality, LncRNA has been found to play a role in a wide range of physiological and pathological processes in carnivores. The comprehensive study of LncRNA not only provides new views on the diagnosis, prevention, and therapy of specific clinical disorders, but also provides fresh insights into the physiological and pathological processes of living creatures. As a result, the study of LncRNA is a highly broad subject with a lot of potential for research [23].

Approximately 98% of human RNA transcripts are non-coding, despite the fact that the mammalian genome generates a large number of transcripts throughout the transcription process. There are several forms of non-coding RNAs with distinct regulatory roles, including microRNAs (miRNAs), long noncoding RNAs (LncRNAs), circular RNAs (circRNAs), and small nucleolar RNAs (snoRNAs). RNAs longer than 200 nucleotides that do not code for proteins are known as LncRNAs. In addition to pathogenic activities like cancer, dysregulation of LncRNAs has been linked to physiological processes including gene expression control, chromatin remodeling, pluripotency maintenance, DNA damage repair, and competing endogenous RNAs. Next-generation sequencing and bioinformatics developments have made it possible to learn more about the roles and activities of LncRNAs, including their ability to encode practical micropeptides. LncRNAs, circRNAs, and pre-miRNAs are examples of ncRNAs that frequently encode functional micropeptides through open reading frames (ORFs). The involvement of LncRNA-encoded peptides in malignant tumors has been the main focus of recent human research on these molecules. These micropeptides, like coding and noncoding genes, are oncogenic drivers or tumor suppressors that play a role in cancer metabolism, tumor angiogenesis, N6-methyladenosine (m6A) modification, and upregulating transmission.

There are several reasons why these peptides might not have been found in earlier research. The physicochemical nature of micropeptides renders conventional mass spectrometry useless, and they are significantly shorter than normal proteins (often less than 100 aa) [24,25,26]. Second, micropeptides are relatively young gene products in human evolution since they are seldom conserved [26, 27, 28]. Third, the bulk of micropeptides are expressed at low quantities, much below the mass spectrometry peptide identification threshold [29]. There is an urgent need for novel techniques to identify these micropeptides, such as using bioinformatics to predict LncRNAs that may encode peptides and then verifying those predictions experimentally.

In GC, some LncRNAs have an oncogenic role. H19 LncRNA 6–7 was shown to be substantially elevated in the tumors of GC patients infected with H. pylori, according to a recent study. In H. pylori-infected GC cells, increased invasion, migration, and proliferation as well as all inflammatory indicators were linked to the overexpression of LncRNA H19. In particular, LncRNA H19 demonstrated good diagnostic accuracy in a Receiver operating characteristic (ROC) study designed to find potential markers for distinguishing stages III to IV of GC (95.5). Jin et al. found that the interstitial fluid of GC patients had a significant increase in the LncRNA HULC, and that overexpression of this LncRNA was associated with H. pylori infection, tumor development, and metastasis. Jing et al. discovered that LncRNA PVT1 expression was significantly greater in normal gastric epithelial cells GES-1 infected with H. pylori. Remarkably, lowering LncRNA PVT1 reduced the inflammatory markers produced by an H. pylori infection. The results showed that H. pylori-associated GC carcinogenesis may be caused by the LncRNA PVT1, which functions as a pro-inflammatory agent. Additionally, it was found that GC patients' poor prognosis, vascular invasion, tumor development, and pathological differentiation are all influenced by aberrant expression of GC-associated LncRNA 1 (GClnc1).

Since overexpressed GClnc1 has been found to be overexpressed in intestinal metaplasia (IM), dysplasia, GC, and normal stomach tissue, it may serve as a diagnostic indicator for early-stage GC. The target gene's histone modification pattern may be altered by GClnc1's interactions with the WDR5 (a key element of the histone methyltransferase complex) and KAT2A histone acetyltransferase complex, which would hasten the development of GC. GClnc1 was similarly shown to be overexpressed in GC tissues infected with H. pylori. However, it remained unclear exactly how GClnc1 contributed to the production of GC caused by H. pylori. However, there is proof that LncRNAs act as tumor suppressors, which has been linked to the development of GC after H. pylori infection. For example, the transcription factor E2F1 11

may be drawn to the promoter of LncRNA AF147447, which interacts with miRNA-34c to regulate the synthesis of MUC2. This suggested that the H. pylori-suppressed LncRNA AF147447 had a role in GC synthesis as a tumor suppressor. Furthermore, It was demonstrated that LncRNA NR 026827 was down-regulated in H. pylori-infected GES-1 cells and that GC tissues expressed less of it than normal tissues did. However, it is unknown how exactly LncRNA NR 026827 contributes to H. pylori-associated GC.

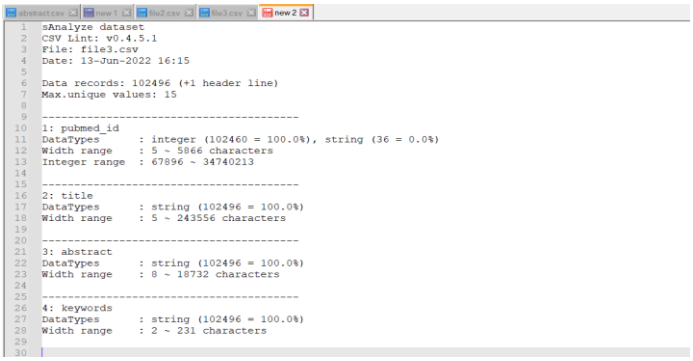
Numerous studies have been conducted on the LncRNA profiles seen in tumors and cells infected with H. pylori. Yang et al. found certain LncRNAs that were abnormally expressed in H. pylori-infected gastric epithelial cells using microarray analysis 13. They discovered that the expression of 23 LncRNAs was upregulated, while the expression of 21 LncRNAs was down regulated. Only the expression levels of XLOC-004122 and XLOC-014388 were found to be decreased in the gastric mucosal samples of patients infected with H. pylori 13, according to further research. In H. pylori-associated GC, Yang et al. found that certain LncRNAs were dysregulated, which may indicate that they had predictive significance 14. Particularly, RP11-169F17.1 and RP11-669N7.2 were shown to be considerably associated with H. pylori infection and to be highly connected with poor overall survival. Zhang et al [30] presents a research on the functions of RNAs and their role in human disease. Their research shows that RNA can be serve as biomarker for curing of various diseases. Chen et al [31] perform a case study on cancer patient to find the association of RNAs with the disease. According to Chen et all (2019) association between RNA and disease can be used to find more complex pathogenesis and to design more effective methods of treatment. Da et al. [32] shows that RNAs is a biomarker in all type of cancers. And it can be used to find and cure cancer diseases

3. Methodology

The procedure of gathering, assessing, and analysing correct insights for research purposes following defined and authorized protocols is known as data collection. Based on the information obtained, a researcher might assess their hypothesis. Regardless of the field of study, the first and most significant phase in the qualitative research is generally data collection. Different methods techniques data collecting is employed in various areas of research based largely on the information necessary. In this study data is collected from an online source known as PubMed shown in figure 2.

Figure 2. PubMed

To implement our model we need an authenticated dataset related to LncRNAs diseases, for this purpose we downloaded data form Pubmed. There are also other sources like GEO or SRA for datasets related to genetic diseases. But in this study we used dataset from Pubmed. It is a well-known authentic source of datasets related to medical field. There is huge amount of data is available on Pubmed related to different diseases association. we can manually download research articles related to any specific disease. It was big challenge for us to download all the research articles related to LncRNA s and its associations with diseases. In this model we required abstracts and keywords from all the articles related to LncRNAs and its association with diseases. We used a list of keywords to search PubMed for published research publications, including LncRNAs, Long non-coding RNAs, LncRNAs in ageing and LncRNAs in senescenes involved in diseases , cancer LncRNAs disease, and LncRNAs in epigenetic diseases , and so on. We acquired more than 90000 publications by retrieving articles and filtering them based on LncRNAs diseases correlations. Figure 3. shows code to analyze keywords of dataset.



```
sAnalyze dataset
CSV Lint: v0.4.5.1
File: file3.csv
Date: 13-Jun-2022 16:15

Data records: 102496 (+1 header line)
Max.unique values: 15

----------------------------------------
1: pubmed_id
DataTypes       : integer (102460 = 100.0%), string (36 = 0.0%)
Width range     : 5 ~ 5866 characters
Integer range   : 67896 ~ 34740213

----------------------------------------
2: title
DataTypes       : string (102496 = 100.0%)
Width range     : 5 ~ 243556 characters

----------------------------------------
3: abstract
DataTypes       : string (102496 = 100.0%)
Width range     : 8 ~ 18732 characters

----------------------------------------
4: keywords
DataTypes       : string (102496 = 100.0%)
Width range     : 2 ~ 231 characters
```

Figure 3. Analyze keywords of dataset

In order to determine their expression or molecular function in controlling target genes or proteins, we focused our data collection efforts on the relationships between LncRNAs and diagnosis. For those LncRNAs, we additionally gathered sequence and location data, followed by experimental procedures, a thorough mechanism and description, in vitro or in vivo investigations, and the reference publication's title and PubMed identifier (ID). We found that the data was in semantic form after compiling it, with lengthy text strings that contained special characters. This might lead to issues when saving and retrieving data from a database. As a result, before saving the data, we used a number of computational pretreatment approaches to ensure that the data could be curated efficiently. Figure 4. shows the proposed scheme.

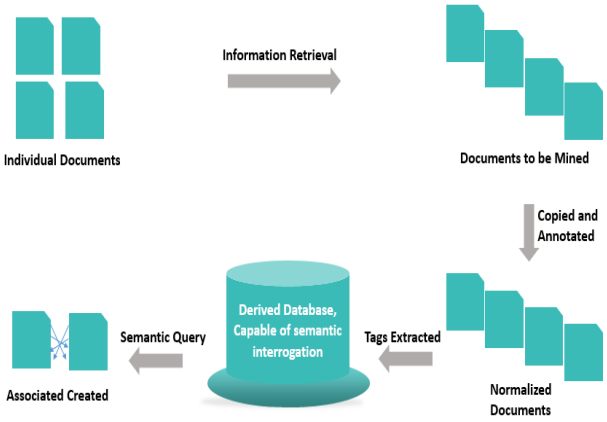


Figure 4. Proposed Scheme

To apply NLP (Natural Language Processing) techniques to complex text data characterizations (such as symbols, double spaces, punctuation, long sentences, grammatical errors) extracted from literature, we have utilized "Natural Language Toolkit" and "TextBlob." We have focused on two areas: "detailed mechanism" and "description." To engage in preprocessing, the following procedures are taken. Figure 5. shows the step involves in our methodology.

3.1. Tokenization

In most cases, the textual description of acquired data from various research publications combines words and incomprehensible symbols, such as special characters and punctuation. When storing data in MYSQL, such symbols create challenges. Tokenization removes superfluous symbols from the text and separates it into tokens[33,34].

3.2 Spelling correction

There may be typographical or spelling mistakes in the unstructured aspects of the data that was gathered, such as the full process and description. Therefore, in this preprocessing phase, we remove such flaws.

3.3 Stop-word removal

In the text of a publication, sentences are usually connected by constructive phrases (such prepositions) and other linguistic structures. Such words are known as stop-words[35.36]. The preprocessed data is free of stop words.

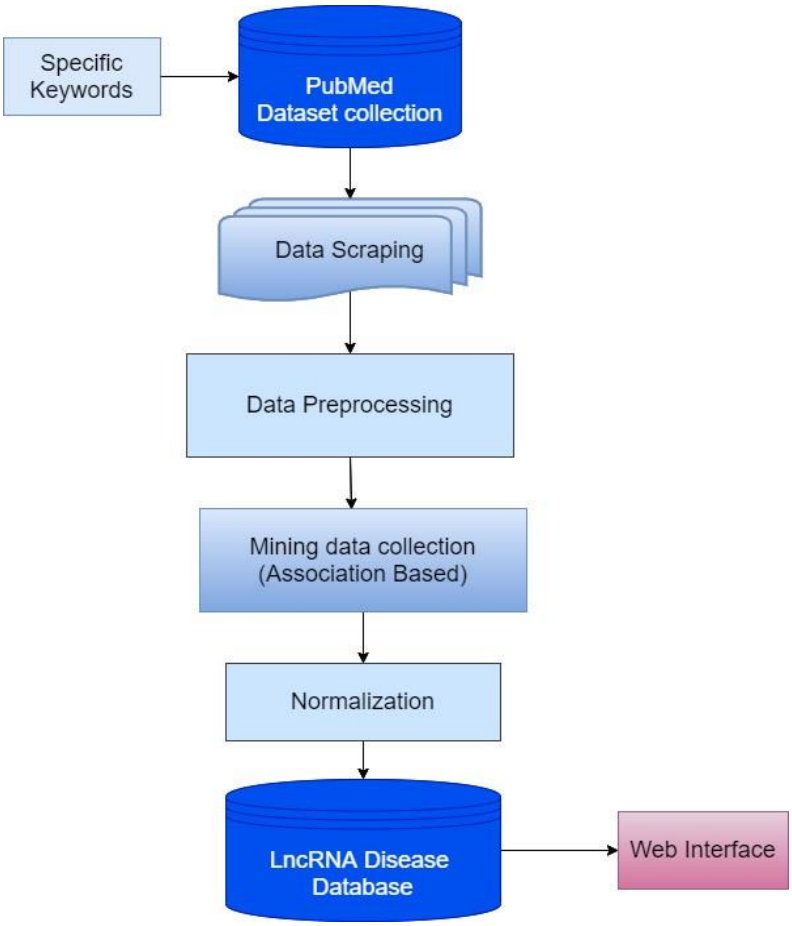


Figure 5. Methodology

3.4 Word inflection and lemmatization

Lemmatization is the process of reducing comparative and superlative phrases to their most basic forms, whereas word inflection reduces words to single forms [37,38]. For instance, lemmatization turns the word "computation" into "compute," while inflection turns the word "bugs" into "bug." We employed both lemmatization and word inflection to get rid of words that were repeated with the same basic concept. The preprocessed words were all altered to lowercase at the end (for instance, "upregulated" to "upregulated"). We selected the illness association data for LncRNAs in the "annotation" field based on experimental considerations of different parameters in the reference studies. Snapshot of dataset after preprocessing is shown as Figure 6.

Figure 6. Snapshot of Dataset

RNA-Seq, whole genome sequencing (WGS), and microarray technologies, for example, identified LncRNAs that were "predicted". The missing sequence and location information for LncRNAs was gathered from LncRNA reference databases (such as nclRNABank and lncRBase) and other non-coding RNA databases (such as NONCODE 3.0). Due to the intricacy of categorization and the several genome versions used by different noncoding RNAs databases in reference research, it was discovered after collecting that the data were highly varied. The information used in the LncRNA-disease association study came from a variety of reference LncRNA databases, each having its own ID. For instance, lnc RNABank and LncRNAQuest may be used to look for certain LncRNAs in reference LncRNA databases and main genome browsers such as GenBank.

In order to enable users to search, explore, and analyze results in the LncRNA Disease database using DQ ID data, for standardization we extracted DQ IDs. Data was normalised before being stored in our database by removing data redundancy and abnormalities. Finally, using MySQL, all of the mined data was put in a database (version 5.7.25). To make the web portal appealing, the web interface was designed in HTML and CSS. The data processing tools were written in PHP (5.7), ajax, and JavaScript, with the web services built on the Xamp server. In conclusion, lncRDisease is a unique database resource that contains 7939 manually curated associations of 4796 experimentally supported LncRNA implicated in 28 different disease types.

User Interface:- You may "search," "browse," and "submit" on the official LncRNA Disease website. A result page for the expression or interaction type of the searched LncRNAs (or disease associated LncRNAs) in that disease will be displayed once users enter the DQ-ID or LncRNA ID and choose a specific disease or any disease to investigate the LncRNAs' association (expression) in that disease Currently, users may peruse LncRNA-disease association data for cancer and epigenetic illnesses. In order to maintain the LncRNA Disease database current, researchers can add new data using the "submit" option.

Additionally, users get access to the "details page," which lists LncRNA target genes along with a thorough explanation of how LncRNA regulation or expression of target genes works. On the basis of experimental methodologies, LncRNAs are classified as anticipated, validated and related in the database's annotation field. The overall functional link is described first, followed by the cells or tissues employed in the reference study. The 'detailed page' also includes the study's LncRNA sequence, species, location, title and PubMed IDs. LncRNA Disease employs a 'non fuzzy' search method to en- sure that a precise match is identified. LncRNAs Disease also includes novel LncRNAs and LncRNA-like RNAs (LncRNA- like), which have been linked to a variety of diseases. For LncRNAs without DQ IDs, LncRNA-like, and novel LncRNAs, LncRNA Disease has its own search ID.

4. Results and Discussion

The spatial and temporal expression of LncRNAs is clearly important for appropriate cellular differentiation and development, from the embryonic stage to gonad development. This makes the dysregulated production of LncRNAs and more precisely, the regulation of their target genes a potentially diagnostic sign for a variety of disorders. Recent research has confirmed the function of LncRNAs in several types of diseases, particularly cancer, Aging and Senescence. In the near future, a significant amount of data pertaining to LncRNA illness associations is anticipated to be generated. Fig 7 and Fig 8 shows the database and table related to LncRNA disease.

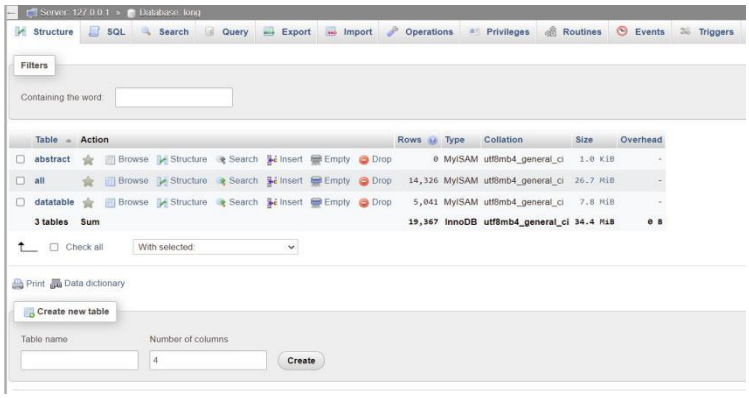


Figure 7. Database

pubmed ID	title	abstract	id
0	title	abstract	0
0	title	abstract	0
0	title	abstract	0
32467232	Role for carbohydrate response element-binding pr...	Long noncoding RNAs (lncRNAs) have been shown to ...	0
33859765	Targeting long noncoding RNA PMIF facilitates ost...	Rationale: The migration of mesenchymal osteoprog...	0
31751568	Transcriptional control of a novel long noncoding...	Differentiated vascular smooth muscle cells (VSMC...	0
32102886	Long Noncoding RNA NRAV Promotes Respiratory Sync...	Respiratory syncytial virus (RSV) is an enveloped...	0
33880577	Long noncoding RNA BANCR promotes proliferation, ...	Esophageal squamous cell carcinoma (ESCC) is a ma...	0
33313940	Long noncoding RNA H19 inhibition ameliorates oxy...	As one of the earliest discovered long noncoding ...	0
32827542	Effect and mechanism of the long noncoding RNA MA...	AIMS: The most typical pathological manifestation...	0
30924864	Long noncoding RNA NEAT1 modulates immune cell fu...	AIMS: Inflammation is a key driver of atheroscler...	0
33901015	Silencing long noncoding RNA NEAT1 alleviates acu...	This study aimed to investigate the role of long ...	0

Figure 8. Database Table

Hence, We created the LncRNAs Disease database by compiling LncRNA-disease relationship data from the literature. LncRNAs Disease is the unique LncRNA database resource, including 97000 LncRNA-disease-related entries, 57096 unique LncRNAs, and 28 categories of associated diseases in human beings. However, the role of LncRNAs in deep regulation mechanisms remains unknown. In human samples of patients, 12 LncRNAs were consistently dysregulated between cases and controls (CCAT1, CRNDE, HOTAIR, CCDC26, LINC00265, SNHG5, KCNQ5IT1, PVT1, MALAT1, TUG1: increased in cases, MEG3 and NEAT1: decreased in cases). We found that a number of malignancies, including those of the breast, stomach, colon, mesothelium, liver, and cervical regions, had dysregulated LncRNAs. Only a small number of studies, nonetheless, have uncovered the precise molecular functions that LncRNAs play in aging in certain disorders. For instance, therapy with androgen and estrogen hormones increased the expression of LncRNA in prostate cancer.

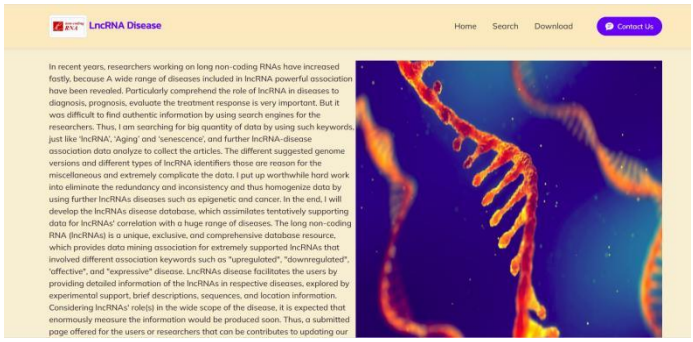


Figure 9. LncRNA Homepage

Furthermore, the overexpression of LncRNA was strongly associated with the growth of the tumor, which was facilitated by the CDK4 and cyclin D1 pathways in "non-small cell lung carcinoma." According to these findings, abnormal expression of "LncRNAs" is important for many cancer kinds, although the precise mechanism behind this is only known for a small number of cancer types. There are now 28 disorders for which LncRNA-disease relationship data are available, of which 54 percent are various types of cancers; 40 percent are cardiovascular diseases; 4 percent are neurodegenerative diseases; and 1 percent are spermatozoa genesis-related and other diseases.

Figure 10. Navigation bar in home page

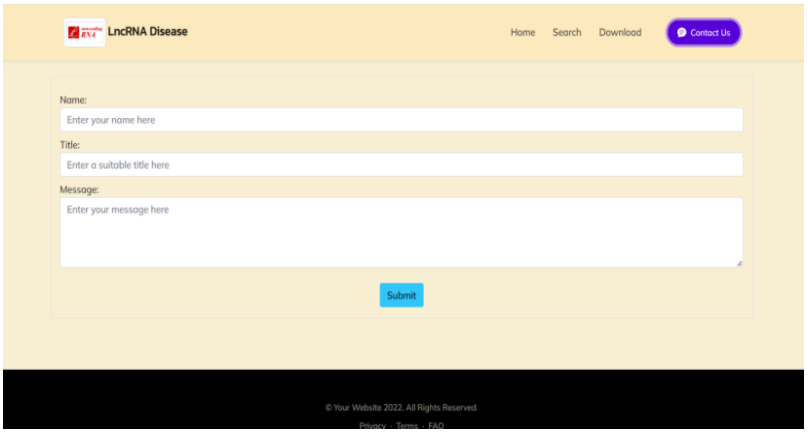


Figure 11. Contact Us Page

Figure 9 shows the homepage of LncRNA and Figure 10. shows the page for searching and downloading the abstracts. In this database we also provide a page to contact with us. Users can submit their queries by using this page. Figure 11. illustrate the page used for the submission of queries.

5. Conclusion

We created the LncRNAs Disease database as a convenient, comprehensive web-based database resource to give a central resource for the biological community to search, analyse, and investigate the LncRNA-disease correlations. Providing thor- ough information on the involvement of LncRNAs in various diseases. LncRNAs Disease benefits the scientific community as a whole.Insights into the functional relationships of LncRNAs across a large range a variety of diseases This new and one- of-a-kind database resource will lead to other research ideas.

References

1. N. Khanna and P. Verma, “microRNA 21 and long non-coding RNAs interplays underlie cancer pathophysiology: A narrative review,” PubMed, 2024.
2. K. Nemeth et al., “Non-Coding RNAs in Human Health and Diseases,” PLoS Genet., vol. 19, no. 3, pp. e1010675, Mar. 2023. doi: 10.1371/journal.pgen.1010675.
3. T. Wu and S. Wang, “Regulatory ncRNAs: miRNAs, siRNAs, piRNAs, LncRNAs and circRNAs,” Trends in Genetics, vol. 39, no. 7, pp. 523–538, Jul. 2024. doi: 10.1016/j.tig.2024.03.005.
4. H. Katsushima and S. Takeuchi, “Potential regulatory roles of microRNAs and long noncoding RNAs in anticancer therapies,” Current Pharmaceutical Design, vol. 25, no. 40, pp. 4293–4305, 2021. doi: 10.2174/1381612825666191021114327.
5. R. Qureshi and M. Mehler, “Long non-coding RNAs: Novel targets for nervous system disease diagnosis and therapy,” Neurotherapeutics, vol. 20, no. 1, pp. 1–15, Jan. 2023. doi: 10.1007/s13311-022-01299-0.
6. I. Yao and M. Mehler, “Long non-coding RNAs in neuronal aging,” Frontiers in Genetics, vol. 10, p. 602736, Feb. 2019. doi: 10.3389/fgene.2019.00602.
7. J. Sun et al., “Long non-coding RNAs: Key regulators of tumor epithelial/mesenchymal plasticity and cancer stemness,” Cancers, vol. 12, no. 5, Art. no. 1245, May 2020. doi: 10.3390/cancers12051245.
8. M. D. Jansson et al., "Regulation of translation by site-specific ribosomal RNA methylation," Nature Structural & Molecular Biology, vol. 28, no. 11, pp. 889-899, 2021.
9. G. C. Cavalcante, L. Magalhães, Â. Ribeiro-dos-Santos, and A. F. Vidal, "Mitochondrial epigenetics: Non-coding RNAs as a novel layer of complexity," International Journal of Molecular Sciences, vol. 21, no. 5, p. 1838, 2020.
10. G. Nigita et al., "ncRNA editing: functional characterization and computational resources," Computational Biology of Non-Coding RNA, pp. 133-174, 2019.
11. W. Berg et al., "Intercalibration of the GPM microwave radiometer constellation," Journal of Atmospheric and Oceanic Technology, vol. 33, no. 12, pp. 2639-2654, 2016.
12. J. X. Yang, R. H. Rastetter, and D. Wilhelm, "Non-coding RNAs: an introduction," Non-coding RNA and the Reproductive System, pp. 13-32, 2016.
13. C. Gao, J. Yang, M. Chen, H. Yan, and X. Wang, "Growth curves and age-related changes in carcass characteristics, organs, serum parameters, and intestinal transporter gene expression in domestic pigeon (Columba livia)," Poultry Science, vol. 95, no. 4, pp. 867-877, 2016.
14. S. Jathar, V. Kumar, J. Srivastava, and V. Tripathi, "Technological developments in LncRNA biology," Long Non Coding RNA Biology, pp. 283-323, 2017.
15. F. Vogel and A. G. Motulsky, Vogel and motulsky's human genetics: Problems and approaches. Springer Science & Business Media, 2013.
16. N. K. Wong, C.-L. Huang, R. Islam, and S. P. Yip, "Long non-coding RNAs in hematological malignancies: translating basic techniques into diagnostic and therapeutic strategies," Journal of Hematology & Oncology, vol. 11, no. 1, pp. 1-22, 2018.
17. L. Lorenzi et al., "The RNA Atlas expands the catalog of human non-coding RNAs," Nature biotechnology, vol. 39, no. 11, pp. 1453-1465, 2021.
18. Y. J. Lee and T. S. Moon, "Design rules of synthetic non-coding RNAs in bacteria," Methods, vol. 143, pp. 58-69, 2018.
19. W. Huanca-Mamani et al., "Long non-coding RNAs responsive to salt and boron stress in the hyper-arid Lluteno maize from Atacama Desert," Genes, vol. 9, no. 3, p. 170, 2018.
20. R. Fujita et al., "Nucleosome destabilization by nuclear non-coding RNAs," Communications biology, vol. 3, no. 1, pp. 1-11, 2020.
21. L. Sun et al., "Transcriptome-wide analysis of pseudouridylation of mRNA and non-coding RNAs in Arabidopsis," Journal of experimental botany, vol. 70, no. 19, pp. 5089-5600, 2019.
22. M. A. Faghihi et al., “Expression of a noncoding RNA is elevated in Alzheimer’s disease and drives rapid feed-forward regulation of beta-secretase,” Nature Medicine, vol. 14, no. 7, pp. 723–730, Jul. 2008. doi: 10.1038/nm1784.
23. S. Jathar, V. Kumar, J. Srivastava, and V. Tripathi, "Technological developments in LncRNA biology," Long Non Coding RNA Biology, pp. 283-323, 2017.
24. W. F. Doolittle, "We simply cannot go on being so vague about ‘function’," Genome biology, vol. 19, no. 1, pp. 1-3, 2018.
25. S. Toden and A. Goel, "Non-coding RNAs as liquid biopsy biomarkers in cancer," British Journal of Cancer, vol. 126, no. 3, pp. 351-360, 2022.
26. L. N. Schulte, W. Bertrams, C. Stielow, and B. Schmeck, "ncRNAs in inflammatory and infectious diseases," Computational Biology of Non-Coding RNA, pp. 3-32, 2019.
27. G. C. Cavalcante, L. Magalhães, Â. Ribeiro-dos-Santos, and A. F. Vidal, "Mitochondrial epigenetics: Non-coding RNAs as a novel layer of complexity," International Journal of Molecular Sciences, vol. 21, no. 5, p. 1838, 2020.
28. G. Nigita et al., "ncRNA editing: functional characterization and computational resources," Computational Biology of Non-Coding RNA, pp. 133-174, 2019.
29. W. Berg et al., "Intercalibration of the GPM microwave radiometer constellation," Journal of Atmospheric and Oceanic Technology, vol. 33, no. 12, pp. 2639-2654, 2016.

30. Y. Zhang, X. Zhang, H. Chen, and Y. Ma, "Circular RNAs: Promising biomarkers for human diseases," Cell Death & Disease, vol. 10, Art. no. 503, Sep. 2019. doi: 10.1038/s41419-019-1720-7
31. X. Chen, J. Wang, M. Li, and H. Liu, "Cancer microRNA association discovery algorithm: A case study on breast cancer," Methods, vol. 160, pp. 123–135, Dec. 2019. doi: 10.1016/j.ymeth.2019.03.009
32. J. Da, Z. Wang, and T. Guo, "Extracellular RNAs as potential biomarkers for cancer," Non-Coding RNA Research, vol. 5, no. 2, pp. 71–80, Apr. 2020. doi: 10.1016/j.ncrna.2020.03.003
33. Javeed, M. U., Shafqat Maria Aslam, Hafiza Ayesha Sadiqa, Ali Raza, Muhammad Munawar Iqbal, & Misbah Akram. (2025). Phishing Website URL Detection Using a Hybrid Machine Learning Approach. Journal of Computing & Biomedical Informatics. Retrieved from https://jcbi.org/index.php/Main/article/view/989.
34. M.U. Javeed, M. S. Ali, A. Iqbal, M. Azhar, S. M. Aslam and I. Shabbir, "Transforming Heart Disease Detection with BERT: Novel Architectures and Fine-Tuning Techniques," 2024 International Conference on Frontiers of Information Technology (FIT), Islamabad, Pakistan, 2024, pp. 1-6, doi: 10.1109/FIT63703.2024.10838424.
35. Javeed, M., Aslam, S., Farhan, M., Aslam, M., & Khan, M. (2023). An Enhanced Predictive Model for Heart Disease Diagnoses Using Machine Learning Algorithms. Technical Journal, 28(04), 64-73. Retrieved from https://tj.uettaxila.edu.pk/index.php/technical-journal/article/view/1828.
36. Aslam, S., Usman Javeed, M. ., Maria Aslam, S. ., Iqbal, M. M., Ahmad, H. ., & Tariq, A. . (2025). Personality Prediction of the Users Based on Tweets through Machine Learning Techniques. Journal of Computing & Biomedical Informatics, 8(02). Retrieved from https://www.jcbi.org/index.php/Main/article/view/796.
37. Raza, A., Zongxin, S., Qiao, G., Javed, M., Bilal, M., Zuberi, H. H., & Mohsin, M. (2025). Automated classification of humpback whale calls in four regions using convolutional neural networks and multi scale deep feature aggregation (MSDFA). Measurement, 255, 118038. https://doi.org/10.1016/j.measurement.2025.118038.
38. Md Neazur Rahman. (2024). Cutting-Edge Novel Method for Credit Card Fraud Detection: Using Data Science Techniques and Machine Learning Algorithms. International Journal of Intelligent Systems and Applications in Engineering, 12(23s), 2040 –. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/7248.