Received: 6 October 2025 | Revised: 7 January 2026 | Accepted: 6 February 2026 | Published online: 26 February 2026

RESEARCH ARTICLE

Medinformatics
2026, Vol. 3(4) 375–386
DOI: 10.47852/bonviewMEDIN62027833

BON VIEW PUBLISHING

Unraveling the Biomarker Prospects of High-Altitude Diseases: Insights from Biomolecular Event Network Constructed Using Text Mining

Balu Bhasuran[1, 2], Sabenabanu Abdulkadhar[3, 4] and Jeyakumar Natarajan[1, 4,*]

[1] *DRDO-BU Center for Life Sciences, Bharathiar University, India*

[2] *School of Information, Florida State University, USA*

[3] *Department of CSE, Koneru Lakshmaiah Education Foundation, India*

[4] *Department of Bioinformatics, Bharathiar University, India*

Abstract: High-altitude diseases (HAD), encompassing acute mountain sickness, high-altitude cerebral edema, and high-altitude pulmonary edema (HAPE), are triggered by hypobaric hypoxia at elevations above 2500 m. These conditions pose significant health risks, yet the molecular mechanisms remain insufficiently understood. In this study, we developed a biomolecular event extraction pipeline integrating supervised machine learning with feature-based and multiscale Laplacian graph kernels to analyze 7847 curated HAD-related abstracts from PubMed. We extracted over 150 unique biomolecular events—including gene expression, regulation, binding, and localization—and constructed a weighted, undirected biomolecular event network comprising 97 nodes and 153 edges. Using the PageRank algorithm, we prioritized key biomolecules based on their centrality within the event network. The top-ranked proteins included erythropoietin (0.0163), vascular endothelial growth factor (0.0148), hypoxia-inducible factor 1 alpha (0.0136), endothelial PAS domain protein 1 and angiotensin-converting enzyme (0.0119), and Egl nine homolog 1, endothelin 1, and 70-kilodalton heat shock protein (0.0118), all of which play crucial roles in oxygen sensing, vascular remodeling, erythropoiesis, and blood pressure regulation. Subnetwork analysis revealed three major functional clusters centered on hypoxia response, inflammation, and stress adaptation pathways. Our integrative approach demonstrates the utility of large-scale text mining and graph-based analysis to uncover mechanistic insights and prioritize potential biomarkers for HAD.

Keywords: biological event extraction, biomarker discovery, natural language processing, high-altitude diseases, text mining

1. Introduction

Altitude sickness, also known as high-altitude illness, is a physiological response to reduced oxygen availability at elevations typically above 2500 m. It manifests in three major clinical forms: acute mountain sickness (AMS), high-altitude cerebral edema (HACE), and high-altitude pulmonary edema (HAPE) [1, 2]. Common symptoms include headache, nausea, loss of appetite, and sleep disturbances [3]. The risk increases with rapid ascent, and millions of individuals—ranging from travelers and athletes to military personnel—are exposed to such environments each year. The onset of high-altitude diseases (HAD) can occur suddenly and may range from mild discomfort to severe, life-threatening conditions [4].

Despite extensive research efforts, much of the valuable biomedical knowledge generated from experimental studies remains locked within the vast and rapidly growing body of scientific literature. Biomedical text mining has emerged as a powerful tool to bridge this knowledge gap by uncovering hidden patterns, generating hypotheses, and identifying novel associations from unstructured text [5–10]. Applications of text mining have shown promise in various domains such as gene-disease associations, drug repurposing, and pharmacogenomics [11–15]. In particular, understanding how genes and proteins behave under hypoxic stress—through events such as transcription, regulation, expression, localization, and binding—offers critical insight into the pathophysiology of HAD [16, 17]. These biomolecular events play a central role in functional genomics, disease mechanisms, and therapeutic targeting [18]. Automatically extracting such events into structured representations, a task known as biomolecular event extraction [19–25], remains technically challenging due to issues

***Corresponding author:** Jeyakumar Natarajan, DRDO-BU Center for Life Sciences, Bharathiar University, India and Department of Bioinformatics, Bharathiar University, India. Email: n.jeyakumar@buc.edu.in

© The Author(s) 2026. Published by BON VIEW PUBLISHING PTE. LTD. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

such as inconsistent terminology, gene/protein name ambiguity, and varying event descriptions [26, 27].

To address these challenges, we developed a supervised machine learning-based text mining system for extracting biomolecular events associated with HAD. Using a curated corpus of PubMed abstracts, our system identified key genes, proteins, and their interactions and constructed a biomolecular interaction network. We then applied the PageRank algorithm to prioritize proteins based on their centrality and connectivity within the network. Our analysis revealed recurring molecular events and highlighted candidate biomarkers potentially involved in the progression of HAD. The remainder of this paper is organized as follows: the Materials and Methods section details the corpus construction and event extraction methodology; the Results section presents the identified biomolecular events and network characteristics; the Discussion interprets key molecular findings; and finally, we conclude with future research directions.

2. Materials and Methods

The biomolecular event extraction pipeline for HAD literature was designed as a multi-step process comprising data collection, preprocessing, event extraction, network construction, network analysis, and biomarker discovery. For data collection, we utilized PubMed (https://pubmed.ncbi.nlm.nih.gov/), a widely used search engine for accessing the MEDLINE database maintained by the National Center for Biotechnology Information, US National Library of Medicine. Relevant HAD literature was retrieved using targeted keyword queries. The resulting records were formatted by extracting the PubMed ID (PMID), title, and abstract for each entry. Preprocessing involved several natural language processing tasks, including sentence splitting, tokenization, part-of-speech (POS) tagging, chunking, and biomedical lemmatization. Syntactic information was then extracted using dependency and phrase structure parsing, followed by the extraction of 14 domain-specific linguistic and semantic features aimed at capturing biomolecular event mentions.

To perform event extraction, we employed a hybrid machine learning approach that integrates a multiscale Laplacian graph (MLG) kernel with a traditional linear feature-based kernel within an ensemble learning framework. This model was used to detect event triggers, while a rule-based pattern-matching engine (using regular expression templates) was implemented to identify event arguments such as themes and causes. The extracted genes/proteins and associated events were used to construct a biomolecular interaction network. We applied graph analysis—specifically the PageRank algorithm—to this network to identify and prioritize key proteins with potential biomarker roles in hypoxia-related pathways. The complete event extraction pipeline is illustrated in Figure 1 and is further elaborated in the sections that follow.

2.1. High-altitude disease information retrieval

In the context of text mining, information retrieval (IR) refers to the process of gathering relevant literature on a specific topic from various biomedical databases. PubMed, a widely used search engine for accessing the MEDLINE bibliographic database—which contains over 34 million citations in the biomedical and life sciences domains—was utilized for this purpose. To compile a comprehensive corpus focused on HAD, we queried PubMed using three key terms: "High-altitude diseases", "High-altitude sickness", and "High-altitude illness". The retrieved articles were downloaded and processed to create a curated HAD-specific text corpus, which included the PubMed ID (PMID), title, and abstract of each article. This corpus served as the foundation for subsequent information extraction and analysis.

2.2. Biomolecular event extraction system

Biomolecular events represent interactions and functional relationships among biological entities such as genes, proteins, and molecules. These events describe key processes—including transcription, regulation, phosphorylation, and binding—that drive cellular and physiological functions. Automatically extracting such events from unstructured biomedical literature and converting them into structured representations is known as event extraction. The resulting event structures are analogous to reactions in biological pathways and are valuable for applications such as pathway curation, biological network construction, and drug repurposing.

In this study, we applied our in-house developed biomolecular event extraction system [28] to a curated corpus on HAD. The system operates in three primary stages: (1) preprocessing and cleaning of textual data, (2) event detection using a hybrid machine learning framework, and (3) event argument identification using pattern-matching techniques. Each step is described in detail below.

2.3. Preprocessing and cleaning

The preprocessing phase involves preparing raw biomedical text for downstream event extraction. This includes sentence segmentation, tokenization, POS tagging, lemmatization, chunking, dependency parsing, and named entity recognition (NER). We used the OpenNLP toolkit (v1.9.0) [29] for sentence segmentation, POS tagging, and chunking, and BioLemmatizer (v1.2) [30] for lemmatization. Dependency parsing was carried out using

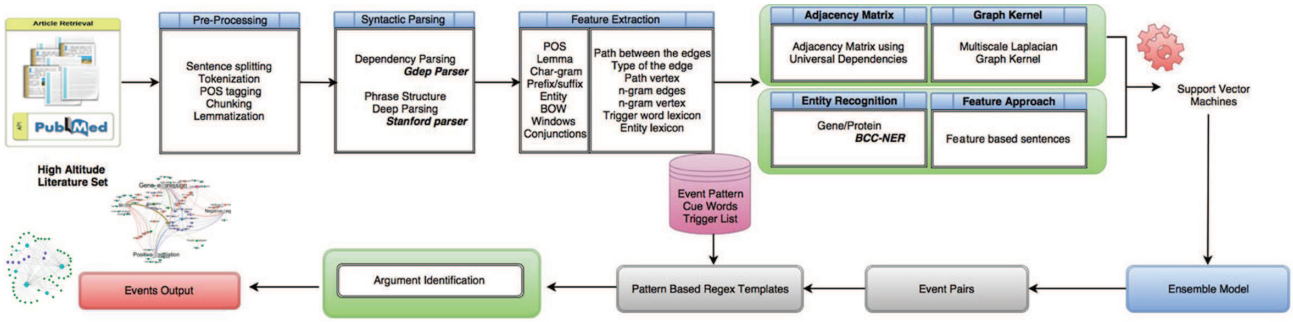


Figure 1. Schematic architecture of the bimolecular event extraction methodology for high-altitude diseases

both the Stanford Parser (v3.3.1) [31] and GDep Parser (v1.0) [32] to identify grammatical structures and syntactic dependencies between trigger words and biological entities. Accurate identification of named entities such as genes and proteins is critical for capturing event participants. For this, we employed BCC-NER [33], our in-house hybrid NER system, which automatically tags biomedical entities relevant to HAD.

2.4. Ensemble classifier

To detect biomolecular events, we implemented a hybrid ensemble learning approach that combines a feature-based kernel with an MLG kernel [34]. The feature-based kernel captures local textual patterns using token-level, sentence-level, and lexical features, while the MLG kernel models complex sentence structures and multilevel topological relationships among event nodes. These two kernels were integrated into an ensemble support vector machine framework [35], allowing the system to leverage both local and global context for robust event identification. Ensemble methods are increasingly popular in biomedical text mining due to their improved generalizability and high classification accuracy [36, 37].

2.5. Event argument detection

The final step involves identifying the roles of entities involved in each event, such as *THEME* and *CAUSE*. We employed a lexico-syntactic pattern-based approach, enhanced with semantic role labeling, to extract arguments accurately from complex biomedical sentences. Lexico-syntactic patterns represent generalized linguistic templates that link trigger words (e.g., activate, bind) with associated biological entities. These patterns were derived through distributional analysis of the training data and refined into rule-based templates for argument detection. Manual rule crafting ensured precise identification of event participants, thereby improving the quality of the extracted event structures. This final structured output was used for downstream network construction and biomarker prioritization.

2.6. Biomolecular event network

The extracted genes, proteins, and associated biomolecular events were integrated to construct a biomolecular event network specific to High-altitude Diseases (HAD). For network construction and visualization, we utilized Gephi version 0.9.2 [38], a widely used platform for interactive graph analysis. The resulting network was modeled as an undirected, weighted graph, where biomolecular events (e.g., gene expression, regulation) functioned as central hub nodes, and edges represented co-occurrences of genes/proteins within those events. Edge weights were assigned based on the frequency of co-occurrence in the HAD corpus, allowing for the representation of interaction strength.

To identify the most influential biomolecules within this network, we applied the PageRank algorithm [39], a node-ranking technique originally developed by Google for ranking web pages. PageRank evaluates the importance of nodes based on both the number and quality of connections they receive. In our context, this allowed us to compute key network attributes—such as PageRank score, weighted in-degree, weighted out-degree, and min-max normalized rank—for each node. These attributes were used to identify critical proteins and to generate subnetworks that highlight the most biologically relevant interactions.

PageRank has been successfully applied in previous biomedical studies for ranking concepts in association networks, identifying research subfields, and re-ranking scientific literature [40–45]. In this study, the algorithm was used to quantify the centrality of each gene/protein in relation to frequently co-occurring events. Nodes with higher PageRank scores were interpreted as more functionally significant in the context of HAD pathogenesis. The network thus becomes a structured map of biomolecular interactions, where hubs (events) and their linked proteins provide insights into biological mechanisms. Inspired by earlier applications, we used PageRank not only to prioritize biomarker candidates but also to construct a normalized subnetwork for deeper analysis of gene-event relationships relevant to high-altitude adaptation and disease. The implementation of PageRank is explained below.

$$\mathrm{PR(u)} = \mathrm{c}\sum_{\mathrm{v}\in\mathrm{B(u)}}\left(\frac{PR(v)}{N_V}\right) \tag{1}$$

Here we have taken u is a node in the network. b(u) is a set of nodes that represent to node *u*, pru and pr(v) are rank scores of the nodes u and v respectively. Nv denotes number of outgoing edges of the node v. c is the factor used to normalize the nodes in the network. Next we have assigned the weights to each node in the network. Then, we measured the weights of inner edges W(v,u)in and outer edges W(v,u)out of the nodes u and v. w(v,u)in is a weight of edge (v,u) calculated based on the number of inner edges of node u and the number of inner edges of v. w(v,u)out is the weight of edge (v,u) calculated based on the number of outer edges of node u and number of outer edges of v. Finally, we employ PageRank to weighted nodes in the network.

$$\mathrm{PR(u)} = (1-\mathrm{d}) + \mathrm{d}\sum_{\mathrm{v}\in\mathrm{B(u)}}\left(\frac{PR(v)\cdot W(v,u)_{in}}{W(v,u)_{out}}\right) \tag{2}$$

Here d is the dampening factor which was used to calculate the probability of edges of each node in the network and (1 – d) is the page rank distribution from non-directly connected nodes.

3. Results

Text mining has proven to be a powerful tool in biomedical research, enabling the discovery of novel biomarkers, uncovering implicit relationships, and generating informative literature-derived networks. These approaches often provide fresh insights into the underlying biological mechanisms of specific diseases [12, 46–48]. In our earlier work, we applied literature mining techniques to investigate gene–gene associations in the context of HAD, revealing functional relationships and shedding light on key genetic interactions involved in hypoxia adaptation [12]. Building on that foundation, the present study focuses on extracting biomolecular events related to HAD using advanced text mining methodologies for the purpose of biomarker discovery. The following results outline the corpus generation process and the findings from our event extraction and network analysis.

3.1. High-altitude disease corpus generation

To construct a comprehensive dataset for event extraction, we performed a targeted keyword search in PubMed up to the year 2024, using three synonymous terms: High-altitude diseases, High-altitude sickness, High-altitude illness. The search results

were downloaded in MEDLINE format as three separate text files. These files were merged and normalized by removing duplicate records, resulting in a nonredundant dataset. Each entry was formatted to include the PubMed ID (PMID), title, and abstract, forming the basis of the HAD corpus. Manual curation was conducted to ensure that all retained articles were directly relevant to HAD research, encompassing both primary studies and reviews. The number of articles retrieved for each keyword query is summarized in Table 1, which forms the foundation for all subsequent biomolecular event extraction and analysis.

Table 1. High-altitude disease dataset downloaded from PubMed using keyword query

Search strategy	Number of records
High-altitude diseases	5422 records
High-altitude sickness	3904 records
High-altitude illness	1023 records
Total	10,349 records

A total of 10,349 unique PubMed records were initially retrieved using the predefined keyword-based search strategies (Table 1) after removal of duplicate PubMed IDs. Following this initial retrieval, inclusion and exclusion criteria were applied to filter out nonrelevant articles. As a result, 7847 records focusing on HAD–related content were retained for advanced analysis and subsequent investigations.

3.2. High-altitude disease event extraction

Studies on HAD data revealed various biological processes involved with genes and proteins in altitude sickness associated with most biomedical events such as expression, regulation, positive, negative regulation, transcription, and binding, etc. The identified events can be used to automatically construct the pathway of altitude sickness.

The following examples explained that the HAD corpus has numerous events, and these events were used to construct the pathway and identify the drug targets.

Phosphorylation played an important role in supplying oxygen to the whole body. In Example 1, proteins, AMP-activated protein kinase, A2B adenosine receptor, BPG mutase, and AMP-activated protein kinase enhanced BPG mutase activation were the most relevant proteins and highly participated in the event of phosphorylation. Our ensemble model efficiently extracts the event phosphorylation and associated proteins to construct the pathway and identify the drug targets in hypoxia.

In Example 2, the proteins angiotensin II receptor 1 gene variants and angiotensin II receptor 1 (AGTR1) were incorporated with the event "binding", and they played a key role in developing the altitude illness HAPE.

In Example 3, the proteins endothelial PAS domain protein 1 (EPAS1) and Egl nine homolog 1 (EGLN 1) were broadly associated with the event binding, and this will help to detect the pathway of AMS. Our ensemble model effectively identified the gene expression and proteins from the corpus by the implementation of MLG and features.

Example: 1

27482003 Mechanistically, we demonstrated that erythrocyteAMP-activated protein kinase was activated in humans at high altitude and that AMP-activated protein kinase is a key protein functioning downstream of the A2B adenosine receptor, phosphorylating and activating BPG mutase and thus inducing 2,3-BPG production and O2 release from erythrocytes. Significantly, preclinical studies demonstrated that activation of AMP-activated protein kinase enhanced BPG mutase activation, 2,3-BPG production, and O2 release capacity in CD73-deficient mice, in erythrocyte-specific A2B adenosine receptor knockouts, and in wild-type mice and in turn reduced tissue hypoxia and inflammation.

T0 ***Protein*** *29 101 Erythrocyte Adenosine A2B Receptor-Mediated AMP-Activated Protein Kinase*
T1 ***Protein*** *365 387 A2B adenosine receptor*
T2 ***Protein*** *424 434 BPG mutase*
T3 ***Protein*** *589 656 AMP-activated protein kinase enhanced BPG mutase activation*
T1038 ***Phosphorylation*** *390 405*
T1039 ***Positive regulation*** *574 584*
E116 ***Phosphorylation:*** *T1038* ***Theme:*** *T0*
E117 ***Phosphorylation:*** *T1038* ***Theme:*** *T1*
E118 ***Phosphorylation:*** *T1038* ***Theme:*** *T2*
E119 ***Positive_regulation:*** *T1039* ***Cause:*** *T3* ***Theme:*** *T3*

Example: 2

27732943 Angiotensin II Receptor 1 gene variants are associated with high-altitude pulmonary edema risk. Previous studies demonstrated that Angiotensin II Receptor 1 (AGTR1) may playan important role in the development of high-altitude pulmonary edema.

T0 ***Protein*** *1 40 Angiotensin II Receptor 1 gene variants*
T1 ***Protein*** *141 166 Angiotensin II Receptor 1*
T1036 ***Binding*** *48 58*
E113 ***Binding:*** *T1036* ***Theme:*** *T0* ***Theme2:*** *T1*
E114 ***Binding:*** *T1036* ***Theme:*** *T1* ***Theme2:*** *T0*

Example: 3

22595196 Direct sequencing was used to identify individualgenotypes for the three SNPs. Age was found to be significantly associated with the EPAS1 SNP in the CMS patients while heart rate (HR) and oxygen saturation level of hemoglobin (SaO(2)) were found to be significantly associated with the EGLN1 (rs480902) SNP in the Han patients with AMS.

T2 ***Protein*** *913 923 EPAS1*
T3 ***Protein*** *999 1009 hemoglobin*
T4 ***Protein*** *1079 1084 EGLN1*
T1003 ***Binding*** *889 899*
T1004 ***Binding*** *1055 1065*
E5 ***Binding:*** *T1004* ***Theme:*** *T3*
E4 ***Binding:*** *T1003* ***Theme:*** *T2*

During the identification of proteins and events associated with HAD, we observed that a specific subset of proteins consistently appeared as both Theme and Cause across multiple instances. This dual role suggests that these proteins play a central role in the underlying biological pathways and may serve as potential biomarkers for HAD. Furthermore, we found that four primary types of events—gene expression, binding, regulation, and localization—frequently co-occur with these proteins. This recurrence indicates that these associations have been substantiated across multiple studies, reinforcing their biological significance.

A comprehensive list of all identified biomolecular events, along with their corresponding PMIDs and abstracts, is provided in Supplementary Data 1. The list of identified gene/protein names along with their associated event theme IDs is available in Supplementary Data 2, and the catalog of biomolecular events annotated with Theme and Event IDs is included in Supplementary Data 3.

3.3. High-altitude disease biomolecular event network

The biomedical event extraction model captures both simple and complex natural language expressions, enabling a wide range of downstream text mining applications such as pathway reconstruction, semantic search, and biomarker discovery [16–28]. In this study, the extracted biomolecular events served as the foundation for constructing a comprehensive network representing the molecular landscape of HAD. This biomolecular event network integrates key genes and proteins that frequently participate in events such as gene expression, regulation, and binding—highlighting their potential roles as biomarkers or therapeutic targets.

The network structure emphasizes the functional interplay between biomolecules and biological processes, with biomolecular events acting as central hub nodes connected to associated gene/protein entities. Figure 2 illustrates this network, offering a systems-level view of the molecular interactions that underpin the pathophysiology of HAD. The network serves not only as a visual representation of literature-derived knowledge but also as a basis for quantitative analysis, including node ranking and subnetwork identification.

A weighted, heterogeneous biomolecular event network was constructed using event themes extracted from the curated HAD literature. In this network, both genes/proteins and biomolecular events (e.g., gene expression, regulation) were represented as nodes. An edge was drawn between a gene/protein and an event node whenever the molecule functioned as a theme within that event. Each edge was assigned a weight proportional to its frequency of occurrence in the HAD corpus—higher co-occurrence frequencies resulted in stronger edge weights, reflecting greater biological relevance. The final network consisted of 97 nodes and 153 edges, capturing a broad spectrum of literature-derived molecular interactions related to HAD. The network depicted in Figure 2 revealed the major proteins, which are acting in these events as vascular endothelial growth factor (VEGF), hypoxia-inducible factor 1(HIF-1) alpha and beta, erythropoietin(EPO), EPAS1, EGLN1, angiotensinogen (AGT), angiotensin-converting enzyme (ACE), endothelin 1 (ET-1), and 70-kilodalton heat shock protein (HSP70). We strongly believe that these proteins play a significant role in HAD and can act as potential biomarkers.

The most frequently occurring event types—gene expression, regulation, binding, and protein catabolism—were further analyzed for their roles in HAD-related pathways and biomarker discovery. Notably, gene expression was linked to conditions such as chronic mountain sickness (CMS) and high-altitude pulmonary hypertension (HAPH). Genes such as HIF, ROS-AMP, and ROE were heavily involved in these events and are known to regulate energy metabolism, catecholamine production, iron homeostasis, vasomotor control, and angiogenesis—key processes in cellular oxygen adaptation. HIF genes, in particular, function as transcriptional regulators and were originally identified in hepatoma cells as inducers of EPO, which plays a protective role against ischemia in CMS.

In binding events, proteins such as EPAS1, EGLN1, and HSP70 were actively involved, revealing their functional roles in the molecular pathways associated with AMS. Meanwhile, regulatory events involving HIF-1α, KOF-2, HSP90, and IL-8 highlighted their significance in controlling hypoxia-related gene activity. Lastly, protein catabolism events featured genes such as HIF, HIF-1α, and EPO, which are potentially implicated in the gastrointestinal complications observed in altitude-related illnesses. The overall distribution and frequency of these biomolecular events across HAD literature are illustrated in Figure 3, offering insight into their potential as drug targets or diagnostic biomarkers.

As shown in Figure 3, gene expression, binding, and positive regulation are the most frequently occurring biomolecular events associated with HAD. To prioritize key molecular players within this event-driven network, we applied the PageRank algorithm, a widely adopted method in network analysis. PageRank assigns a numerical weight to each node—such as a gene or protein—based on its connectivity and influence within the network. This measure reflects not only the number of connections a node has but also the significance of the nodes it is connected to.

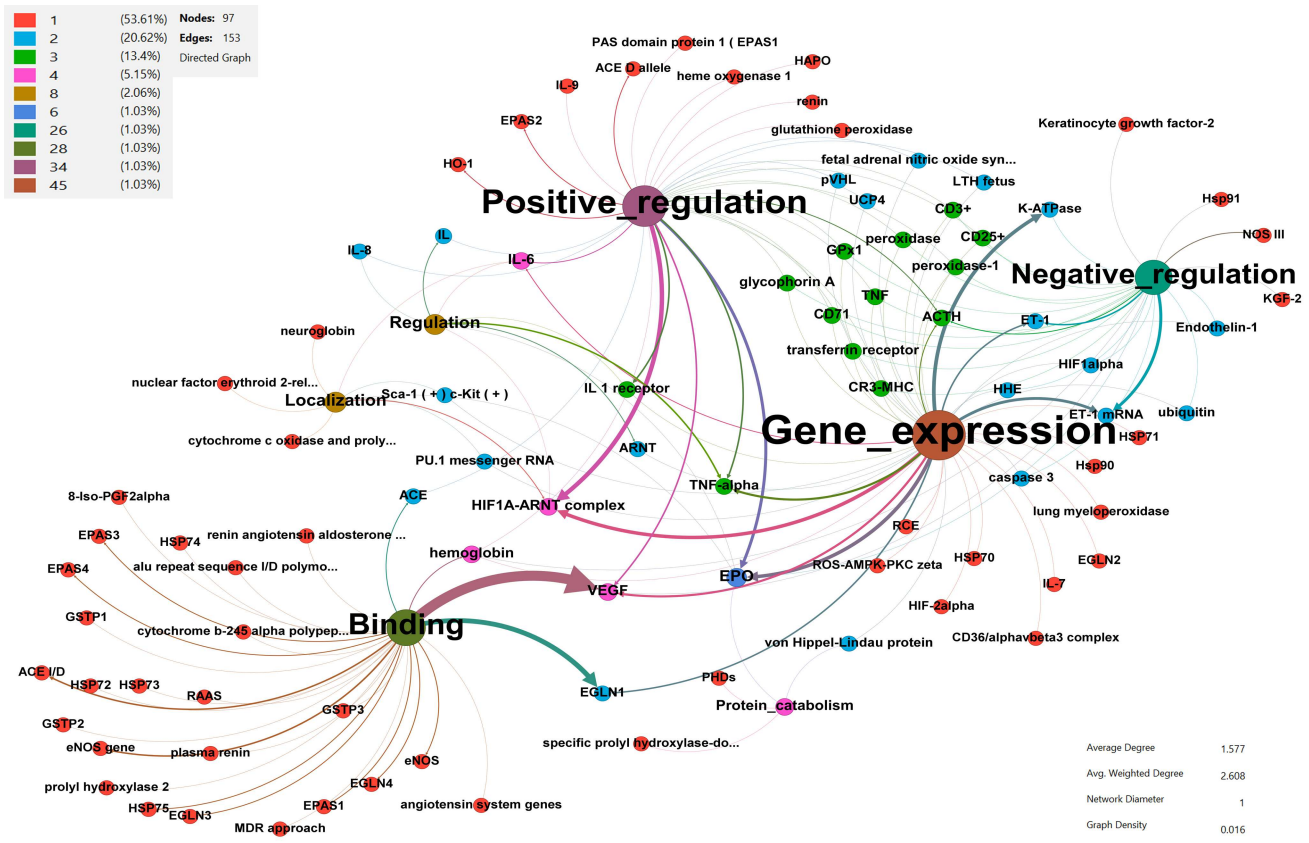


Figure 2. Biomolecular event network for high-altitude disease illustrating key molecular entities and biological processes involved in the response to hypoxic conditions

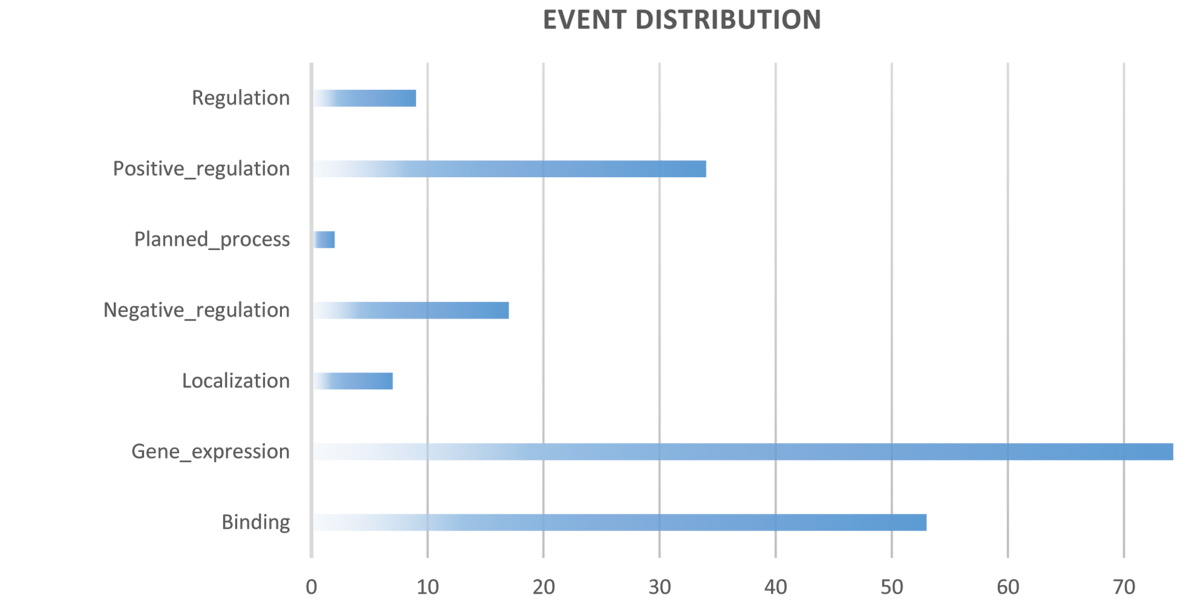


Figure 3. Distribution of biomolecular events in high-altitude diseases

The PageRank analysis revealed several high-impact genes, including EPO (0.0163), VEGF (0.0148), HIF-1 (0.0136), EPAS1 and ACE (0.0119), and EGLN1, HSP70, and ET-1 (0.0118). These proteins emerged as central entities within the network due to their frequent and diverse involvement in key events. Their prominent ranks underscore their functional relevance in hypoxia adaptation, vascular regulation, and erythropoiesis—critical biological responses in HAD. The full results of the PageRank analysis, including numerical rankings and connectivity scores, are provided in Supplementary Data 4, and the full

biomolecular event network file is available as Supplementary Data 5.

By leveraging the PageRank algorithm, we identified gene expression, regulation, binding, and protein catabolism as event hubs that orchestrate the molecular pathways involved in HAD. The corresponding high-ranking proteins—EPO, VEGF, HIF-1, HSP70, ACE, EGLN1, and ET-1—can be considered potential biomarkers or therapeutic targets. Their network attributes, such as degree, proximity, and normalized min–max rank, are summarized in Figure 4, while Figure 5 presents the distribution of key network metrics, further validating their central roles in the HAD biomolecular landscape.

Figure 6 shows the subnetwork of complex molecular interactions underlying HAD, emphasizing hypoxia-driven regulatory pathways. The network is organized into three major clusters: Cluster 1 encompasses inflammatory cytokines such as IL-6 and IL-8, which modulate immune and localization responses; Cluster 2 centers on hypoxia-inducible genes such as HIF1α, ET-1, and endothelin-1, which contribute to vascular adaptation through negative gene regulation; and Cluster 3 involves stress-related and iron-regulating factors such as ACTH, TNF, and the transferrin receptor. Key processes such as positive regulation and gene expression serve as central hubs linking hypoxia signaling (e.g., HIF1A-ARNT complex) to downstream effectors such as EPO and VEGF, highlighting their role in erythropoiesis and vascular remodeling. This systems-level representation underscores the interplay of immune, endocrine, and transcriptional responses critical to physiological adaptation and pathology at high altitude. We searched for biological databases for the major genes/proteins and the connected biological pathways, and provided substantial evidence in Table 2.

Table 2 summarizes the major genes and proteins identified through biomolecular event extraction and network analysis, along with their associated biological pathways relevant to HAD. Key pathways include the HIF-1 signaling pathway, which features central hypoxia-responsive genes such as EPO, VEGFA, HIF1A, EPAS1, and EGLN1, all of which are critical for oxygen sensing, erythropoiesis, and vascular remodeling under hypoxic conditions. The renin-angiotensin system (RAS) pathway includes AGT, ACE, and AGTR1, implicating blood pressure regulation in HAD pathophysiology. The vascular smooth muscle contraction pathway involves genes such as MAPK1 and EDNRA, indicating roles in vasoconstriction and vascular tone. Additionally, genes such as IL6, IL8, and TNF are involved in inflammatory and immune responses, while HSP70 and HSP90 represent key stress-related proteins contributing to cellular protection mechanisms. This comprehensive mapping of genes to pathways highlights the multifactorial nature of HAD and underscores the potential of these biomolecules as diagnostic or therapeutic targets.

Finally, we investigated the functional roles of some of these genes and identified that EGLN1 is an oxygen sensor catalyzer, EPAS1 plays a role in the rise to the blood-brain barrier, VEGFA induces the blood vessels permeabilization, and AGT is a potent blood pressure regulator [49–51]. Since HAD has been related to a lack of oxygen and hypertension, the roles of these genes are vital in the exploration of future studies.

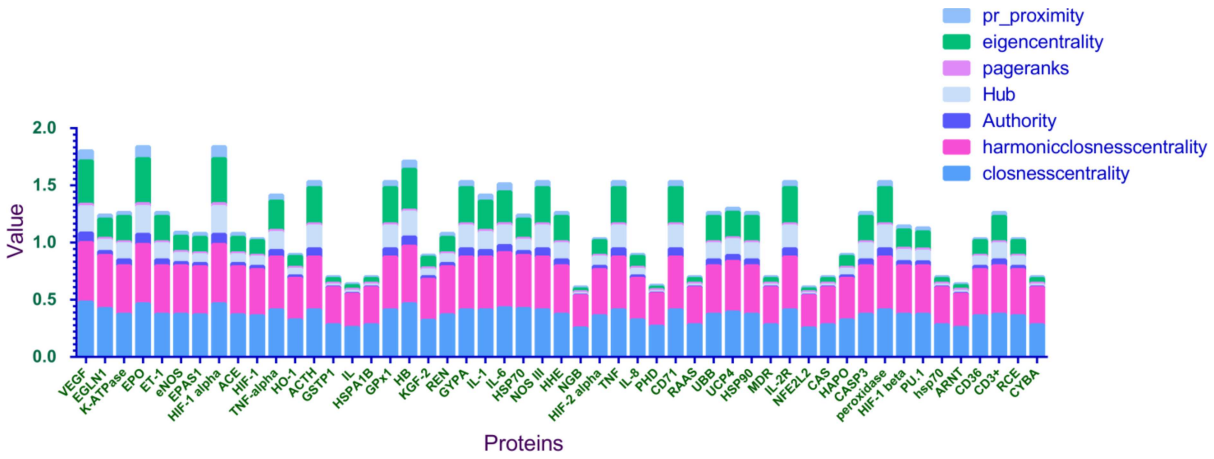


Figure 4. Top genes/proteins in biomolecular events of high-altitude diseases based on the degree and PageRank within the network

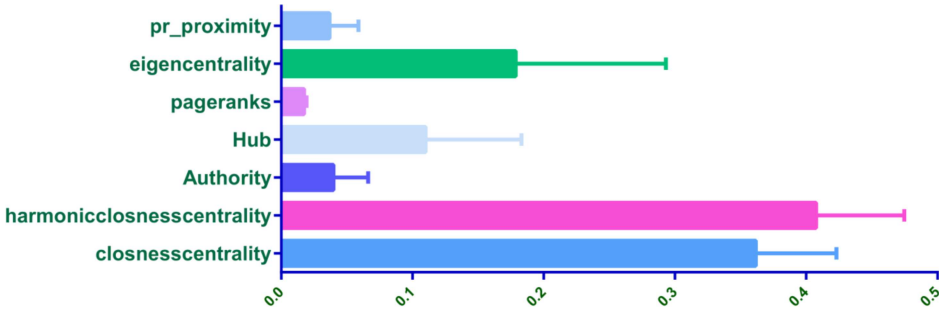


Figure 5. Distribution of the various network statistical measures applied to the biomolecular event network of high-altitude diseases

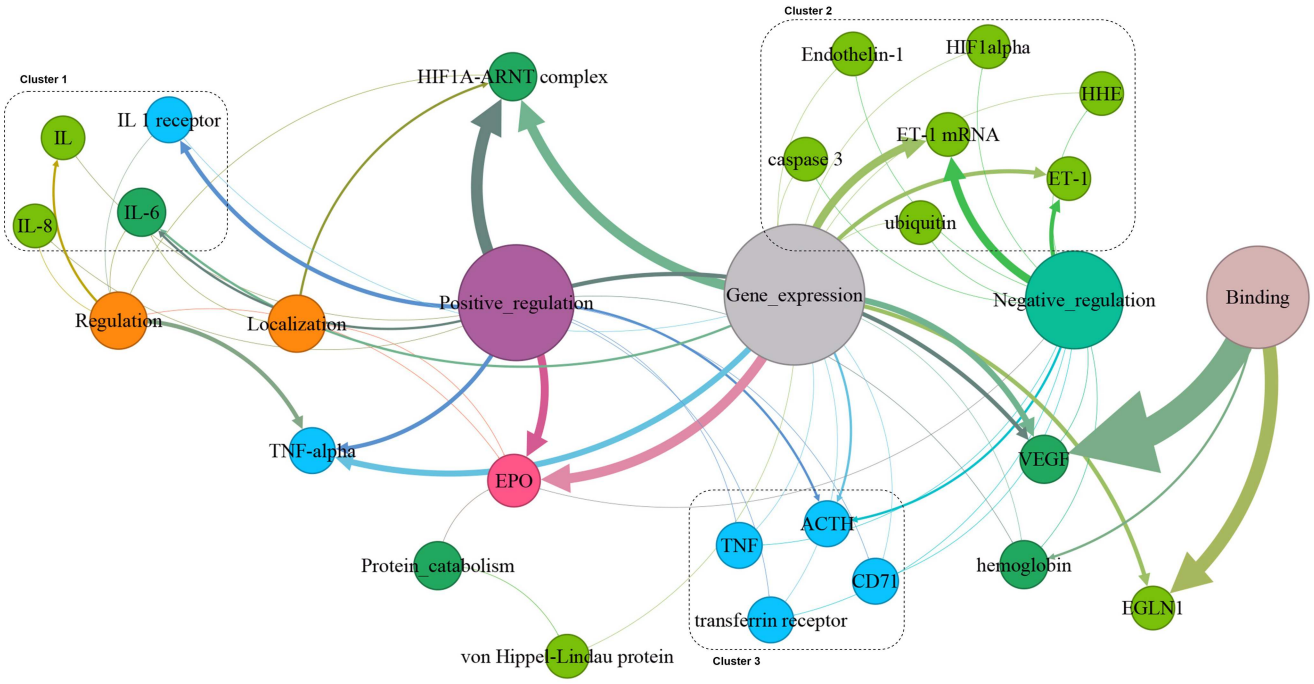


Figure 6. Biomolecular event subnetwork for high-altitude disease illustrating key molecular entities and biological processes involved in the response to hypoxic conditions

Table 2. Major genes/proteins and the connected biological pathways in high-altitude diseases

Pathway	Genes/proteins	Biological database
HIF-1 signaling pathway (hsa04066)	EPO (erythropoietin), VEGFA (vascular endothelial growth factor A), HIF1A (hypoxia-inducible factor 1 alpha), EPAS1, LDHA, NOS3, PHD1, EDN1, AKT3, ANGPT1	KEGG Pathway Reactome (e.g., "Signaling by HIF") WikiPathways (e.g., WP1249 "HIF-1 signaling") PANTHER Pathways
Renin-angiotensin system (RAS) (hsa04614)	AGT (angiotensinogen), AGTR1, AGTR2, ACE (angiotensin-converting enzyme), REN	KEGG Pathway Reactome (e.g., "Renin-angiotensin system") WikiPathways (e.g., WP1801 "Renin-Angiotensin System") PANTHER Pathways
Vascular smooth muscle contraction (hsa04270)	AGTR1, EDNRA (endothelin receptor type A), MAPK1	KEGG Pathway Reactome (e.g., portions under "Smooth muscle contraction") WikiPathways (e.g., WP1384 "Vascular smooth muscle contraction") PANTHER Pathways
Hypoxia response and iron regulation	EGLN1, HIF1A, EPAS1, TFRC (transferrin receptor), LDHA	Reactome "Cellular response to hypoxia" "Iron uptake and transport" WikiPathways (e.g., WP1445 "Hypoxia response"; WP133 "Iron metabolism") Pathway Commons PANTHER (under "Hypoxia response" and "Iron homeostasis")
Inflammatory signaling and localization	IL6, IL8, TNF, ACTH	Reactome (e.g., "Cytokine signaling in immune system") KEGG Pathway (e.g., "Cytokine–cytokine receptor interaction" hsa04060) WikiPathways (e.g., WP492 "Cytokine signaling in immune response") PANTHER Pathways ("Inflammation mediated by chemokine and cytokine signaling pathway")

(*Continued*)

Table 2. (*Continued*)

Pathway	Genes/proteins	Biological database
Protein folding and stress response	HSP70, HSP90	Reactome (e.g., "Chaperonin-mediated protein folding"; "HSF1 activation") WikiPathways (e.g., WP2806 "Heat shock response") PANTHER Pathways ("Heat shock response") KEGG Pathway ("Protein processing in endoplasmic reticulum" hsa04141 includes chaperones)

4. Discussion

HAD, including AMS, HACE, and HAPE, represent a multifactorial physiological challenge involving oxygen deprivation, vascular dysregulation, inflammation, and stress response. Despite the growing number of clinical and molecular studies, there remains a critical gap in systematically identifying key biomolecules and pathways associated with HAD. In this study, we demonstrated the value of biomedical text mining as a scalable strategy to extract biomolecular events from literature, enabling a data-driven approach to biomarker discovery. By integrating NER, syntactic parsing, and a hybrid ensemble classification model, we identified high-confidence biological events such as gene expression, regulation, binding, and localization from over 7800 PubMed abstracts related to HAD. These events offer a granular representation of how genes and proteins interact during hypoxic stress, and when organized into a network structure, they offer deeper insight into mechanistic relationships. PageRank-based analysis added an interpretive layer, ranking biomolecules by their connectivity and influence within the event network, thus aiding in prioritizing biomarker candidates for further investigation.

The biomolecular event network (Figure 2) constructed from the extracted events consisted of 97 nodes and 153 edges, where nodes represent either biomolecular entities or events and edges capture co-occurrence relationships derived from the literature. Events such as gene expression, binding, positive/negative regulation, and protein catabolism emerged as central hubs, reflecting their frequent and functionally diverse involvement in HAD. Proteins with the highest PageRank scores—such as EPO, VEGF, HIF-1α, EPAS1, EGLN1, ACE, ET-1, and HSP70—demonstrated widespread connectivity across multiple event types. This suggests their systemic involvement in physiological adaptations to hypoxia, including erythropoiesis, angiogenesis, and cellular stress response. Notably, the repeated association of these proteins with multiple event types (as both Theme and Cause) strengthens their biological relevance and implies their potential utility as biomarkers or therapeutic targets. For instance, the prominent role of HIF-1 [52–54] and its downstream targets such as EPO [55] and VEGF [56] aligns well with established knowledge on hypoxia signaling and high-altitude adaptation, thereby validating the model's ability to recover biologically significant relationships from unstructured text. To further refine our understanding, we performed a subnetwork analysis (Figure 6) that illuminated clustered patterns of molecular interaction within the main network. This analysis revealed three functional clusters: the first featured inflammatory mediators such as IL-6 [57] and IL-8 [58, 59], which are known to modulate immune activation and vascular permeability in high-altitude exposure; the second was centered on hypoxia-inducible regulators such as HIF1α [52–54], EGLN1 [60–62], and EPAS1 [60–62], reflecting their established role in vascular homeostasis and oxygen sensing; and the third included stress-related and metabolic regulators such as ACTH [63, 64], TNF [63, 64], and transferrin receptor, highlighting links to neuroendocrine signaling and iron metabolism. Central hubs in the subnetwork—positive regulation, binding, and gene expression—linked these clusters, showing how a relatively small set of recurring event types mediate diverse physiological processes in HAD. The subnetwork layout also highlighted the bidirectional and multi-event participation of top-ranked genes, underscoring their integrative role across pathways. These findings reinforce the utility of subnetwork analysis in dissecting complex molecular relationships and offer a focused map of high-value biomarker targets for future experimental validation.

This study demonstrates the effectiveness of combining biomedical text mining with network-based analysis to uncover key biomolecular players and regulatory mechanisms in HAD. By systematically extracting and analyzing biomolecular events from literature, we constructed a comprehensive event network and identified central hub proteins using the PageRank algorithm. These high-confidence proteins—such as EPO, VEGF, HIF-1α, EPAS1, EGLN1, ACE, ACTH, TNF, ET-1, IL-6, IL-8, and HSP70—are strongly implicated in hypoxia adaptation and offer promising leads for further investigation as biomarkers or therapeutic targets. Moving forward, the integration of additional data sources such as gene expression profiles from high-altitude exposed populations, proteomics, or metabolomics datasets could significantly enhance the network's biological richness and predictive power. Incorporating such multi-omics evidence into the existing literature-derived network will enable more precise identification of disease modules, dynamic signaling cascades, and context-specific biomarkers, ultimately facilitating translational research and personalized intervention strategies in altitude-related pathophysiology.

5. Conclusion

Text mining offers powerful means to uncover hidden associations and novel patterns within the biomedical literature. In this study, we built a specialized corpus of 7847 PubMed records to identify key genes and proteins involved in HAD. By applying the BCC-NER tool for gene/protein recognition and a hybrid machine learning classifier (combining feature-based and MLG kernels) for event extraction, we assembled a biomolecular interaction network and used PageRank to highlight the most influential nodes. Our analysis revealed that hypoxia-related events—particularly gene expression, binding, regulation, and localization—occur repeatedly in the context of HAD. Central genes such as VEGFA, HIF1A, EPO, EPAS1, and EGLN1 emerged as critical hubs in hypoxia signaling, while AGT, ACE, and ET-1 connected hypoxic responses to blood pressure regulation. HSP70 also featured prominently, reflecting its role in stress adaptation under low-oxygen conditions. Together, these biomolecular events and their associated

genes underscore the intricate cross-talk between oxygen sensing, vascular homeostasis, and stress response pathways in HAD. Further investigation of these hypoxia-driven events is likely to yield valuable biomarkers and therapeutic targets for mitigating the physiological challenges of high-altitude exposure.

Acknowledgment

This work was supported by the DRDO-BU Centre for Life Sciences, Coimbatore, Tamil Nadu, India. The author (Balu Bhasura) acknowledges the fellowship received from the grant.

The authors acknowledge that an earlier version of this manuscript was published as a preprint at arXiv (DOI: https://doi.org/10.48550/arXiv.2507.10953) on July 15, 2025. This published article incorporates revisions made during the peer review process for *Medinformatics*.

Funding Support

This work is sponsored by the funding from DRDO-BU Centre for Life Sciences, Bharathiar University, Coimbatore-641046, Tamil Nadu, India.

Ethical Statement

This study did not require ethical approval because it was conducted entirely using computational methods and publicly available data/resources, without involving human participants, animal subjects, or identifiable personal data. Therefore, no human or animal experiments were performed by any of the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Balu Bhasuran: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Sabenabanu Abdulkadhar:** Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Jeyakumar Natarajan:** Conceptualization, Supervision, Project administration, Funding acquisition.

Supplementary Information

The supplementary file is available at https://doi.org/10.47852/bonviewMEDIN62027833.

References

[1] Gatterer, H., Villafuerte, F. C., Ulrich, S., Bhandari, S. S., Keyes, L. E., & Burtscher, M. (2024). Altitude illnesses. *Nature Reviews Disease Primers*, *10*(1), 43. https://doi.org/10.1038/s41572-024-00526-w

[2] Richalet, J.-P., Hermand, E., & Lhuissier, F. J. (2024). Cardiovascular physiology and pathophysiology at high altitude. *Nature Reviews Cardiology*, *21*(2), 75–88. https://doi.org/10.1038/s41569-023-00924-9

[3] Li, X., Zhang, J., Liu, G., Wu, G., Wang, R., & Zhang, J. (2024). High altitude hypoxia and oxidative stress: The new hope brought by free radical scavengers. *Life Sciences*, *336*, 122319. https://doi.org/10.1016/j.lfs.2023.122319

[4] Walsh, B., & Agrawal, S. (2024). High-altitude cerebral edema and high-altitude pulmonary edema diagnosed in the desert: A case report. *Clinical Practice and Cases in Emergency Medicine*, *8*(3), 202–205. https://doi.org/10.5811/cpcem.3851

[5] Bhasuran, B. (2022). Combining literature mining and machine learning for predicting biomedical discoveries. In K. Raja (Ed.), *Biomedical text mining* (pp. 123–140). Springer. https://doi.org/10.1007/978-1-0716-2305-3_7

[6] Cheerkoot-Jalim, S., & Khedo, K. K. (2021). A systematic review of text mining approaches applied to various application areas in the biomedical domain. *Journal of Knowledge Management*, *25*(3), 642–668. https://doi.org/10.1108/jkm-09-2019-0524

[7] Naseem, U., Khushi, M., Khan, S. K., Shaukat, K., & Moni, M. A. (2021). A comparative analysis of active learning for biomedical text mining. *Applied System Innovation*, *4*(1), 23. https://doi.org/10.3390/asi4010023

[8] Bhasuran, B., & Natarajan, J. (2023). DisGeReExT: A knowledge discovery system for exploration of disease-gene associations through large-scale literature-wide analysis study. *Knowledge and Information Systems*, *65*(8), 3463–3487. https://doi.org/10.1007/s10115-023-01862-1

[9] Percha, B. (2021). Modern clinical text mining: A guide and review. *Annual Review of Biomedical Data Science*, *4*, 165–187. https://doi.org/10.1146/annurev-biodatasci-030421-030931

[10] Chen, Q., Sun, H., Liu, H., Jiang, Y., Ran, T., Jin, X., ..., & Niu, Z. (2023). An extensive benchmark study on biomedical text generation and mining with ChatGPT. *Bioinformatics*, *39*(9), btad557. https://doi.org/10.1093/bioinformatics/btad557

[11] Bhasuran, B., & Natarajan, J. (2019). Distant supervision for large-scale extraction of gene-disease associations from literature using DeepDive. In *International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018, 2*, 367–374. https://doi.org/10.1007/978-981-13-2354-6_39

[12] Bhasuran, B., Subramanian, D., & Natarajan, J. (2018). Text mining and network analysis to find functional associations of genes in high altitude diseases. *Computational Biology and Chemistry*, *75*, 101–110. https://doi.org/10.1016/j.compbiolchem.2018.05.002

[13] Zhao, S., Su, C., Lu, Z., & Wang, F. (2021). Recent advances in biomedical literature mining. *Briefings in Bioinformatics*, *22*(3), bbaa057. https://doi.org/10.1093/bib/bbaa057

[14] Luo, R., Sun, L., Xia, Y., Qin, T., Zhang, S., Poon, H., & Liu, T.-Y. (2022). BioGPT: Generative pre-trained transformer for biomedical text generation and mining. *Briefings in Bioinformatics*, *23*(6), bbac409. https://doi.org/10.1093/bib/bbac409

[15] Wei, C.-H., Allot, A., Lai, P.-T., Leaman, R., Tian, S., Luo, L., ..., & Lu, Z. (2024). PubTator 3.0: An AI-powered literature resource for unlocking biomedical knowledge.

Nucleic Acids Research, *52*(W1), W540–W546. https://doi.org/10.1093/nar/gkae235

[16] Wu, P., Li, X., Gu, J., Qian, L., & Zhou, G. (2024). Pipelined biomedical event extraction rivaling joint learning. *Methods*, *226*, 9–18. https://doi.org/10.1016/j.ymeth.2024.04.003

[17] Liu, L., Liu, M., Liu, S., & Ding, K. (2024). Event extraction as machine reading comprehension with question-context bridging. *Knowledge-Based Systems*, *299*, 112041. https://doi.org/10.1016/j.knosys.2024.112041

[18] Frisoni, G., Moro, G., & Carbonaro, A. (2021). A survey on event extraction for natural language understanding: Riding the biomedical literature wave. *IEEE Access*, *9*, 160721–160757. https://doi.org/10.1109/ACCESS.2021.3130956

[19] Fei, H., Ren, Y., Zhang, Y., Ji, D., & Liang, X. (2021). Enriching contextualized language models from knowledge graphs for biomedical information extraction. *Briefings in Bioinformatics*, *22*(3), bbaa110. https://doi.org/10.1093/bib/bbaa110

[20] Bali, M., & Anandaraj, S. P. (2023). Biomolecular event extraction using natural language processing. *International Journal of Electrical and Computer Engineering Systems*, *14*(5), 601–612. https://doi.org/10.32985/ijeces.14.5.12

[21] McClosky, D., Riedel, S., Surdeanu, M., McCallum, A., & Manning, C. D. (2012). Combining joint models for biomedical event extraction. *BMC Bioinformatics*, *13*(11), S9. https://doi.org/10.1186/1471-2105-13-S11-S9

[22] Björne, J., & Salakoski, T. (2011). Generalizing biomedical event extraction. In *Proceedings of BioNLP Shared Task 2011 Workshop*, 183–191.

[23] Zhao, W., Zhang, J., Yang, J., He, T., Ma, H., & Li, Z. (2021). A novel joint biomedical event extraction framework via two-level modeling of documents. *Information Sciences*, *550*, 27–40. https://doi.org/10.1016/j.ins.2020.10.047

[24] Chen, Y. (2021). A transfer learning model with multi-source domains for biomedical event trigger extraction. *BMC Genomics*, *22*(1), 31. https://doi.org/10.1186/s12864-020-07315-1

[25] Li, Q., Li, J., Sheng, J., Cui, S., Wu, J., Hei, Y., . . . , & Yu, P. S. (2024). A survey on deep learning event extraction: Approaches and applications. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(5), 6301–6321. https://doi.org/10.1109/TNNLS.2022.3213168

[26] Lai, V. D., van Nguyen, M., Kaufman, H., & Nguyen, T. H. (2021). Event extraction from historical texts: A new dataset for black rebellions. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 2390–2400. https://doi.org/10.18653/v1/2021.findings-acl.211

[27] Hakala, K., van Landeghem, S., Salakoski, T., van de Peer, Y., & Ginter, F. (2013). EVEX in ST'13: Application of a large-scale text mining resource to event extraction and network construction. In *Proceedings of the BioNLP Shared Task 2013 Workshop*, 26–34.

[28] Abdulkadhar, S., Bhasuran, B., & Natarajan, J. (2021). Multiscale Laplacian graph kernel combined with lexico-syntactic patterns for biomedical event extraction from literature. *Knowledge and Information Systems*, *63*(1), 143–173. https://doi.org/10.1007/s10115-020-01514-8

[29] Kwartler, T. (2017). *Text mining in practice with R*. Wiely.

[30] Liu, H., Christiansen, T., Baumgartner, W. A., & Verspoor, K. (2012). BioLemmatizer: A lemmatization tool for morphological processing of biomedical text. *Journal of Biomedical Semantics*, *3*(1), 3. https://doi.org/10.1186/2041-1480-3-3

[31] Socher, R., Bauer, J., Manning, C. D., & Ng, A. Y. (2013). Parsing with compositional vector grammars. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 455–465.

[32] Sagae, K., & Tsujii, J.-i. (2010). Dependency parsing and domain adaptation with data-driven LR models and parser ensembles. In H. Bunt, P. Merlo, & J. Nivre (Eds.), *Trends in parsing technology: Dependency parsing, domain adaptation, and deep parsing* (pp. 57–68). Springer. https://doi.org/10.1007/978-90-481-9352-3_4

[33] Murugesan, G., Abdulkadhar, S., Bhasuran, B., & Natarajan, J. (2017). BCC-NER: Bidirectional, contextual clues named entity tagger for gene/protein mention recognition. *EURASIP Journal on Bioinformatics and Systems Biology*, *2017*(1), 7. https://doi.org/10.1186/s13637-017-0060-6

[34] Kondor, R., & Pan, H. (2016). The multiscale Laplacian graph kernel. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2990–2998.

[35] Claesen, M., de Smet, F., Suykens, J. A. K., & de Moor, B. (2014). EnsembleSVM: A library for ensemble learning using support vector machines. *Journal of Machine Learning Research*, *15*, 141–145.

[36] Chung, J., & Teo, J. (2023). Single classifier vs ensemble machine learning approaches for mental health prediction. *Brain Informatics*, *10*(1), 1. https://doi.org/10.1186/s40708-022-00180-6

[37] Ganie, S. M., & Malik, M. B. (2022). An ensemble machine learning approach for predicting Type-II diabetes mellitus based on lifestyle indicators. *Healthcare Analytics*, *2*, 100092. https://doi.org/10.1016/j.health.2022.100092

[38] Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media*, *3*(1), 361–362. https://doi.org/10.1609/icwsm.v3i1.13937

[39] Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The pagerank citation ranking: Bring order to the web. In *Proceedings of ASIS, 98,* 161–172.

[40] Zeng, Y., Wei, Z., Pan, Z., Lu, Y., & Yang, Y. (2022). A robust and scalable graph neural network for accurate single-cell classification. *Briefings in Bioinformatics*, *23*(2), bbab570. https://doi.org/10.1093/bib/bbab570

[41] Aftiss, A., Lamsiyah, S., El Alaoui, S. O., & Schommer, C. (2024). BioMDSum: An effective hybrid biomedical multi-document summarization method based on PageRank and Longformer encoder-decoder. *IEEE Access*, *12*, 188013–188031. https://doi.org/10.1109/access.2024.3514915

[42] Gupta, S., Sharaff, A., & Nagwani, N. K. (2023). Graph ranked clustering based biomedical text summarization using top-k similarity. *Computer Systems Science and Engineering*, *45*(3), 2333–2349. https://doi.org/10.32604/csse.2023.030385

[43] Ma, K., Li, J., Zhao, M., Zamit, I., Lin, B., Guo, F., & Tang, J. (2024). PPRTGI: A personalized PageRank graph neural network for TF-target gene interaction detection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *21*(3), 480–491. https://doi.org/10.1109/TCBB.2024.3374430

[44] Deng, J. (2025). Enhancing peer review in educational assessments via social media platforms: A fuzzy-CDF and weighted PageRank approach. *Journal of Computational*

Methods in Sciences and Engineering, *25*(2), 1303–1315. https://doi.org/10.1177/14727978241299538

[45] Zhang, Z., Jia, Y., Zhang, X., Papadopoulou, M., & Roche, C. (2023). Weighted co-occurrence bio-term graph for unsupervised word sense disambiguation in the biomedical domain. *IEEE Access*, *11*, 45761–45773. https://doi.org/10.1109/access.2023.3272056

[46] Lima, T., Ferreira, R., Freitas, M., Henrique, R., Vitorino, R., & Fardilha, M. (2022). Integration of automatic text mining and genomic and proteomic analysis to unravel prostate cancer biomarkers. *Journal of Proteome Research*, *21*(2), 447–458. https://doi.org/10.1021/acs.jproteome.1c00763

[47] Li, X., Dai, A., Tran, R., & Wang, J. (2023). Identifying miRNA biomarkers for breast cancer and ovarian cancer: A text mining perspective. *Breast Cancer Research and Treatment*, *201*(1), 5–14. https://doi.org/10.1007/s10549-023-06996-y

[48] Lever, J., Jones, M. R., Danos, A. M., Krysiak, K., Bonakdar, M., Grewal, J. K., . . . , & Jones, S. J. (2019). Text-mining clinically relevant cancer biomarkers for curation into the CIViC database. *Genome Medicine*, *11*(1), 78. https://doi.org/10.1186/s13073-019-0686-y

[49] Wang, S., Hong, Q., Zheng, Y., Duan, S., Cai, G., & Chen, X. (2024). Epigenetics in high-altitude adaptation and disease. *Science Bulletin*, *69*(24), 3806–3811. https://doi.org/10.1016/j.scib.2024.11.008

[50] Chen, Y., Jiang, C., Luo, Y., Liu, F., & Gao, Y. (2016). Interaction of CARD14, SENP1 and VEGFA polymorphisms on susceptibility to high altitude polycythemia in the Han Chinese population at the Qinghai-Tibetan Plateau. *Blood Cells, Molecules, and Diseases*, *57*, 13–22. https://doi.org/10.1016/j.bcmd.2015.11.005

[51] Bhattacharyya, P., Deb, P., & Pradhan, D. (2023). Genetics in high altitude medicine. In J. Hidalgo, S. Da Re, & A. G. D'Almeida (Eds.), *High altitude medicine: A case-based approach* (pp. 23–29). Springer. https://doi.org/10.1007/978-3-031-35092-4_3

[52] Yfantis, A., Mylonis, I., Chachami, G., Nikolaidis, M., Amoutzias, G. D., Paraskeva, E., & Simos, G. (2023). Transcriptional response to hypoxia: The role of HIF-1-associated co-regulators. *Cells*, *12*(5), 798. https://doi.org/10.3390/cells12050798

[53] Lyu, Y., Yang, Y., Talwar, V., Lu, H., Chen, C., Salman, S., . . . , & Semenza, G. L. (2024). Hypoxia-inducible factor 1 recruits FACT and RNF20/40 to mediate histone ubiquitination and transcriptional activation of target genes. *Cell Reports*, *43*(4), 113972. https://doi.org/10.1016/j.celrep.2024.113972

[54] Koyasu, S., Horita, S., Saito, K., Kobayashi, M., Ishikita, H., Chow, C. C., . . . , & Harada, H. (2023). ZBTB2 links p53 deficiency to HIF-1-mediated hypoxia signaling to promote cancer aggressiveness. *EMBO Reports*, *24*(1), EMBR202154042. https://doi.org/10.15252/embr.202154042

[55] Alqarni, A. A., Aldhahir, A. M., Alghamdi, S. A., Alqahtani, J. S., Siraj, R. A., Alwafi, H., . . . , & Pang, L. (2023). Role of prostanoids, nitric oxide and endothelin pathways in pulmonary hypertension due to COPD. *Frontiers in Medicine*, *10*, 1275684. https://doi.org/10.3389/fmed.2023.1275684

[56] Droma, Y., Hanaoka, M., Kinjo, T., Kobayashi, N., Yasuo, M., Kitaguchi, Y., & Ota, M. (2022). The blunted vascular endothelial growth factor-A (VEGF-A) response to high-altitude hypoxia and genetic variants in the promoter region of the *VEGFA* gene in Sherpa highlanders. *PeerJ*, *10*, e13893. https://doi.org/10.7717/peerj.13893

[57] Huang, H., & Yang, L. (2023). Research progress of inflammatory factors in chronic obstructive pulmonary disease with pulmonary hypertension at high altitude. *Alternative Therapies in Health and Medicine*, *29*(8), 689–693.

[58] Karakurum, M., Shreeniwas, R., Chen, J., Pinsky, D., Yan, S. D., Anderson, M., . . . , & Stern, D. (1994). Hypoxic induction of interleukin-8 gene expression in human endothelial cells. *The Journal of Clinical Investigation*, *93*(4), 1564–1570. https://doi.org/10.1172/JCI117135

[59] Wang, Z.-Y., Liu, Y., Li, S.-P., Li, J.-J., Zhang, Z., Xiao, X.-C., . . . , & Yang, S. (2022). Hypoxia inducible factor 1α promotes interleukin-1 receptor antagonist expression during hepatic ischemia-reperfusion injury. *World Journal of Gastroenterology*, *28*(38), 5573–5588. http://dx.doi.org/10.3748/wjg.v28.i38.5573

[60] Gou, Q., & Xue, H. (2025). Association study of EPAS1 gene polymorphisms with the susceptibility to high altitude pulmonary hypertension in Tibetan residents. *Journal of Hypertension*, *43*(Suppl. 1), e203. https://doi.org/10.1097/01.hjh.0001117280.31257.2e

[61] Dai, Z. (2024). Translational potential of hypoxia-inducible factor-2α signaling in pulmonary hypertension. In Y. Zhao (Ed.), *Lung biology and pathophysiology* (1st ed.) (pp. 87–100). CRC Press. https://doi.org/10.1201/9781003355243-6

[62] Tashi, Q.-Z., Tsering, S.-B., Zhou, N.-N., Zhang, Y., Huang, Y.-J., Jia, J., & Li, T.-J. (2022). A study on the molecular mechanism of high altitude heart disease in children. *Pharmacogenomics and Personalized Medicine*, *15*, 721–731. https://doi.org/10.2147/PGPM.S356206

[63] Zangeneh, F. Z., & Shoushtari, M. S. (2022). The significant role of the sympathetic nervous system in hospitalization and death of patients with COVID-19: Psycho-neuroendocrine-immune aspect of stress. *Acta Medica Iranica*, *60*(9), 532–548. https://doi.org/10.18502/acta.v60i9.11094

[64] Pena, E., El Alam, S., Siques, P., & Brito, J. (2022). Oxidative stress and diseases associated with high-altitude exposure. *Antioxidants*, *11*(2), 267. https://doi.org/10.3390/antiox11020267

How to Cite: Bhasuran, B., Abdulkadhar, S., & Natarajan, J. (2026). Unraveling the Biomarker Prospects of High-Altitude Diseases: Insights from Biomolecular Event Network Constructed Using Text Mining. *Medinformatics*, *3*(4), 375–386. https://doi.org/10.47852/bonviewMEDIN62027833