Improved Data-Driven Insights Analysis of Women's CEAP Varicose Vein Risk Factors

[1]S. Chitra Nayagam, [2]Dr. C. Priya

Research Scholar
[1]Department of Computer Applications
Dr.M.G.R. Educational and Research Institute, Chennai, TamilNadu,
chitranayagam77@gmail.com

[2]Professor and Research Supervisor
Department of Computer Applications
Dr.M.G.R. Educational and Research Institute, Chennai, TamilNadu
drcpriya.research@gmail.com

KEYWORDS

data analytics, varicose veins, VVs, genomic, demographical factors, venous.

ABSTRACT

Varicose veins are swollen, twisted, and enlarged veins that are visible just under the surface of the skin commonly found on the legs and feet. They persist due to poorly functioning valves in the veins that leads to massive pooling of blood on the legs. Identifying the demographic and genetic factors leading to varicose veins could help in diagnosing it at the early stages and erstwhile be prevented ahead of its existence. A detailed inclination of both demographic and genetic factors has been interpreted in this study that leads to varicose veins. In addition to the demo-graphical factors like lifestyle, age, pregnancy count, weight, profession etc., certain reports have revealed the contribution and intensity of genetic factors that had led to varicose veins. Details from Genome Wide Association Study have prompted the accountability of genetic factors leading to the prevalence of varicose veins. The study highlighted the association between the genomics position and certain specific chromosomes leading to the traits of varicose veins. Predominant research studies on considering the demo-graphical factors have shown a prominent impact of varicose veins existence among women gender as against their counterparts.

I. INTRODUCTION

Data Analytic thrives to bring more insights by enhancing and furnishing precise and accurate processes and services for any applications. Considering the rapidly growing health related problems among the humans, predominantly for women, a detailed study is required to be conducted to identify the various common ailments that persist among them. The facts and study reports from various research studies have confirmed that the inclining IT techniques could support analyzing, any form of structured or semi-structured or unstructured data by applying data analytics can derive meaningful information using Data Mining techniques and could provide contemporary solutions to various ailments occurring among women. Big data analytics analyzes the entire data set and derives new insights or patterns that exists among the various components and leads to more accurate and relevant information as compared to the sampling techniques. The data analytical tools available in the recent times are more powerful and rapidly render useful information which is in high demand especially in the field of healthcare. The impacts of certain peculiar diseases in women are more as compared to men since the health condition depends on the body anatomy and physiology. There are some common health issues which have intense effect on women as the factors causing them could be associated with the age of women, pregnancy, gender specific physiology, hormonal changes. Different studies related to Varicose Veins have inadvertently proven that women are highly prone to the disorder in comparison with their counterpart.

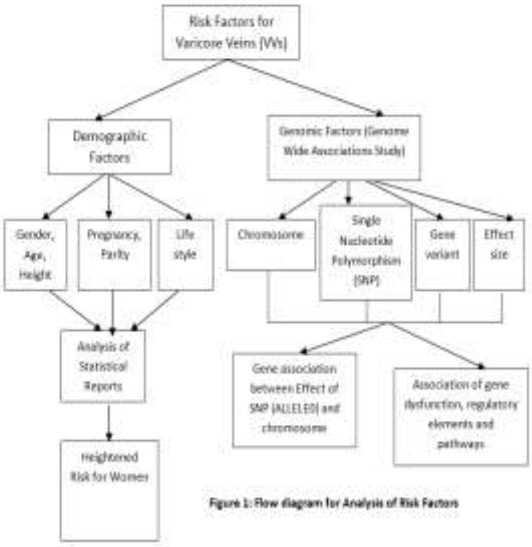


Figure 1: Flow diagram for Analysis of Risk Factors

Varicose Veins (VVs) are swollen, twisted, and enlarged veins that are visible just under the surface of the skin. It is a prevalent venous (blood vessel) disorder and identifying the hindrance indicative factors are very crucial. Varicose veins occur when the valves in the veins that regulate blood flow become weak or damaged. Normally, these valves help blood flow back to the heart, but when they malfunction, blood can pool in the veins, causing them to stretch, enlarge, and appear twisted. The severity of VVs could even lead to Mitral valve regurgitation. It is a condition in which the mitral valve in the heart does not close tightly, causing blood to flow backward (regurgitate) into the left atrium instead of moving forward into the left ventricle and out to the body. This back flow can strain the heart and can lead to heart attack (myocardial infarction) caused by damaged muscles that control the valve.

The demographic factors like gender, age, height, weight, life style, education, profession etc. have aggravated impact on Varicose Veins [2]. Figure 1 depicts the flow diagram for the analysis of various risk factors concerned in the research work. The various studies undertaken from different parts of the world have revealed the effects of demographic factors for varicose veins.

In addition, further factors like excessive weight, aging, and pregnancy count are the leading contributors to the onset of varicose veins. These elements can cause increased pressure in the leg veins appearing it to be dark purple or blue in colour, making the veins to become bulging and twisting. This would lead to itching or skin irritation causing muscle cramping or even bleeding around the affected area.

The research studies have also revealed that the genetic factors are also the sources of VV risk up to 90%. In Western nations, VVs are found in 10%–15% of males and 20%– 25% of females [1].

The historical data consisting of similar symptoms and indications that are collected through various methods could be normalized and made consistent to provide meaningful results. The implications of data analytics on collected data could bring in novel perceptions that might prevent VVs or could predict the complications of VVs in the early stage.

This research work is focused to discuss the impact of demographic factors as well as analyzing the genetic factors that could lead to varicose veins. The paper is organized as section II summarizing the Literature study, section III reviewing Genome Wide Association Study on Kaggle dataset for varicose veins, section IV aggregating GWAS dataset, section V with future work and section VI leading to the conclusion of research.

II. REVIEW OF LITERATURE

The severe effects and the major causes for VVs have been discussed in [2]. As per the CEAP (Clinical-Etiology-Anatomy-Pathophysiology) classification of chronic venous diseases, VVs has taken its position as C2 (class 2) which is third category among eleven in the list. Later CEAP has introduced another subclass C2r which is mentioned as Recurrent VVs. The effects and the causes of VVs leading to various other complications have been discussed in the study.

The saphenous veins are the blood vessels present on the legs that carry deoxygenated blood, also known as venous blood, from the feet and legs back to the heart which can get damaged causing Varicose Veins (VVs). Another factor for VVs, are the perforate veins which are responsible for connecting superficial and deep venous systems which can forbid the blood flow backward thus can lead to muscles contraction. The various demographic risk factors that lead to the weak or damaged valves causing VVs include the gender being women, age above 50, number of pregnancy, hereditary, obesity, deep vein thrombosis (DVT), menopause etc. The other generic factors leading to VVs includes profession that needs standing or sitting posture for longer duration, hypertension, diabetes, physical damage to legs. The literature study regarding the global impact and contributing factors leading to VVs have considered research material that incorporates global data and had presented a precise statistical report on it [5]. This includes Epidemiology which revealed the mechanism of getting affected by VVs based on the various risk factors like age, gender, family history, occupation, pregnancy and parity, obesity and hereditary. The potential risk factors leading to VVs in female victims include pregnancy and parity where the risk rate increases substantially for the women who had more number of deliveries. VVs are more realizable during early stages of pregnancy due to over secretion of progesterone.

As per the study of [1], VVs were found on women with 50.5% as against men with 30.1%. Apparently, venous symptoms leading to VVs were found to be 51.3% in women where as men endures only about 20.4%. This study has been undertaken without any ultrasound examination but with sheer evaluation by trained vascular medical expertise through medical examination. Moreover, VVs are classified as saphenous where the diseased have at least one damaged vein on one lower limb and practically they will be smaller in size and another category being non-saphenous where the damaged vein would be present either on one or both the lower limbs. Statistical analyses have reported that the women were significantly affected with VVs where height and pregnancy have utmost association with the risk factors. Considerable variations of VVs among men and women were detected through the results drawn from the study in case of non-saphenous varicose veins. The research method used a stratified sampling technique that has performed χ^2tests (chi-square) which uses large sample sets for categorical data to identify any association using SPSS software.

The experience of discomfort among patients with venous disorders differs across various countries. Age emerges as the primary risk factor for both genders, with around fifty percent of women and one-third of men being impacted. Notably, elevated prevalence rates have been observed in industrialized nations. Additionally, lifestyle modifications have resulted in reduced disparities in the environmental factors influencing both men and women. It is essential to verify any demographic factors that may be relevant for the causal of VVs [4].

The relationship among varicose veins, skin trophic changes and venous symptoms in both women and men were analyzed. The study had highlighted the significant correlation among varicose veins, ankle pitting edema, and trophic changes, despite the latter two being present in only a minor fraction of individuals with varicose veins. Additionally, the relationship between pitting edema and venous symptoms is evident, with a notable overlap that were highly pronounced in women further persists between varicose veins and venous symptoms. The prevalence of venous symptoms in relation to age reveals a significant negative correlation in women ($P < .001$) even after accounting for the presence of varicose veins. Conversely, in men, the apparent significant increase in venous symptoms with age is negated when adjusted for varicose veins. The study among 33 individuals conducted to identify the significant associations of the risk factors for venous symptoms exhibiting pitting ankle edema, only one did not report venous symptoms (P < .001), rendering it impractical to include this variable in the logistic regression analysis. Female gender had an intense risk factor which was witnessed by a considerably higher prevalence as compared to men [4].

The paper titled [2], Global impact and contributing factors in varicose vein disease development have elucidated certain facts regarding the risk factors leading to VVs. The primary risk factors that may lead to the development and exacerbation of varicose veins (VVs) includes age, gender, family medical history, occupation, genetic predisposition, obesity, physical activity, smoking habits, number of pregnancies, body mass index (BMI), and current lifestyle choices. The research paper has discussed an epidemiological study conducted in the United States revealed that among the individuals who are aged between 40 to 80 years, approximately 11 million men and 22 million women are affected by varicose veins. The paper has concluded that further research is necessary to elucidate the connections between the different risk factors, such as family history, smoking, parity, and body mass index (BMI) that may exacerbate the condition of varicose veins (VVs). Additionally, it is essential to identify the molecular targets implicated in the progression of VVs.

A nationwide population-based retrospective cohort study [3] was conducted utilizing medical claims data from the Taiwan National Health Insurance Research Database (NHIRD) has been depicted in the research work, where Varicose Veins are associated with an increased risk of mitral valve regurgitation (MR). Mitral valve regurgitation is a cardiac condition characterized by the improper closure of the mitral valve, which results in the backward flow of blood into the left atrium. The paper has highlighted the research gap regarding the better analysis on semi-structured data which may reveal associations between varicose veins (VV) and mitral regurgitation (MR). It is advisable to utilize alternative methods for assessing disease severity, including peripheral duplex ultrasound for VV and echocardiography for MR.

III. **REVIEW OF GENOME-WIDE ASSOCIATION STUDY (GWAS)**

The proposed research study has victimized the kaggle dataset done by GWAS that comprises of genetic factors leading to varicose veins using statistical techniques [7]. The Dataset is a summary statistics for discovery GWAS of 22,473 varicose veins cases and 379,183 non-varicose veins controls in UK Biobank [7]

The parameters for the study includes SNP : Name or identifier of the single nucleotide polymorphism (SNP); CHR : Chromosome number where the SNP is present; BP - Base pair genomic position of the SNP on the chromosome; ALLELE1 : (derives from Greek prefix allele -, which means "mutual", "reciprocal", or "each other), an effect allele which is associated with the trait or condition being studied, is a variant of a gene, or a different version of a DNA sequence at a specific location on a chromosome; ALLELE0 : a non effect ALLELE which is reference allele or alternate allele and the possible data for both ALLELE1 and ALLELE0 refers to one of the data in the list {A,T, C or G}. The other genetic elements are A1FREQ: is the frequency of the effect ALLELE in the overall study population and the value ranges from 0 to 1; INFO: represents the imputation quality score for SNPs and again the values for this factor ranges from 0 to 1. The decimal value represents the percentage of the effect ALLELE present in the overall population; BETA which represents the effect size of the ALLELE for the VVs trait.

A) **Association between Chromosomes and Genomics Locations in Varicose Veins:**

A chromosome is a long, continuous thread of DNA that contains many genes, regulatory elements, and other nucleotide sequences. Every organism possesses certain number of chromosomes which represents their specific characteristics. Humans have 23 pairs of chromosomes (which is 46 in total). Chromosomes serve as the structural units of the genome. The genomics position refers to the specific location of a nucleotide, gene, or feature on a chromosome. It is typically expressed in terms of the chromosome number (e.g., "chr1" for chromosome 1). A numerical coordinate of a chromosome specifies the base-pair position within that chromosome. For example, chr1:1234567 indicate a position on chromosome 1 at base-pair coordinate 1,234,567. Coordinates are typically based on a reference genome assembly, such as GRCh38 for humans.

The GWAS study [7] has confirmed that the chromosomes serve as the fundamental structures of DNA that contain genes. The genomic locations specify the precise sites of genes or variants, including single nucleotide Polymorphisms (SNPs), on these chromosomes. In relation to varicose veins, genetic researches, including Genome-Wide Association Studies (GWAS), have revealed that a particular genomics region (loci) is linked to a heightened risk of developing varicose veins.

B) Genetic Findings Related to Varicose Veins

GWAS research has identified several chromosomes and genomics positions associated with the development of varicose veins. For example, variants found near the FOXC2 gene, which is crucial for regulating vein structure and the development of valves, have been linked to this condition. Another recent study [6] had addressed a specific SNP on chromosome 16, such as chr16:53420341 (this is a fictitious example meant for clarity) where chromosomes and genomics positions have been linked to varicose veins. The genes located at these positions are often involved in important functions like maintaining vein wall elasticity, regulating blood flow, and facilitating the proper development of venous valves.

C) **Mechanism related to Gene Dysfunction, regulatory elements and pathways**

Genes located at these genomics position could be responsible for producing proteins that help sustain the structure and function of veins. Any mutations or variations at these sites might compromise the integrity of vein walls or hinder valve function, which could result in blood pooling and lead to vein enlargement.

Variants in the regulatory regions of DNA, located adjacent to genes, can impact gene expression, thereby affecting the health of veins.

Genes found in related regions generally tend to get involved in the development and remodelling of blood vessels, structure of connective tissues, including the synthesis of collagen and elastin and also formation of valves in veins.

Genome-Wide Association Study has identified genetic variations that are linked to varicose veins by scanning the entire genomes of affected and unaffected individuals. The study had further identified the high-risk genomics positions that could help to predict the individuals who might develop varicose veins. This could help early intervention of the risk involved in varicose veins.

IV. AGGREGATION OF GWAS DATA SET:

A) Preprocessing genomic Dataset

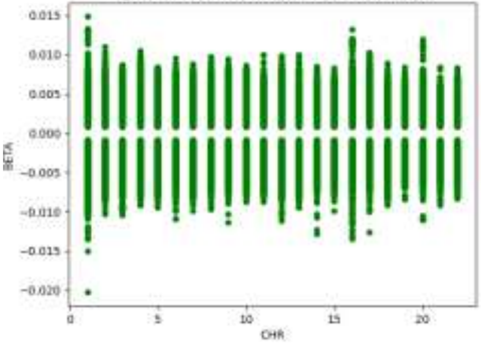
The big data set considered for the study had very large data points with SNP, CHR, ALLELE1, ALLELE0, A1FREQ, PVAL and BETA as the dimensions. The data set published by GWAS was in text format and therefore pre-processing step was included before proceeding further with the study.

Initially, the text data has been converted to csv file where columns are separated and then made it ready for processing. The ALLELE1 and ALLELE0 factors in the original data set is assigned one of the characters among A, T, C, G which is then converted to numerical values 1,2,3,4 respectively. Later this numerical data has been utilized for statistical techniques to visualize the associations among the parameters.

B) **Analysis between CHR and BETA**

The association between chromosome number (CHR) where the SNP is present and BETA which is the effect size of the SNP present for the trait (varicose veins) is depicted in Figure 2.

Figure 2 : Chromosome Vs BETA (Effect of SNP)



The chromosome numbers varies from 1 to 22 and BETA value which ranges from 0 to 1 on both sides of the number scale represents the effect size of a specific genetic variant (SNP) on the trait of varicose veins. It quantifies the association between the allele of the SNP and VVs, typically under a linear or logistic regression model. The positive value of BETA signifies the intense effect of the allele that could increase the risk of VVs and the negative value has less impact on VVs. Figure 2 depicts that the allele had large impact on chromosome 1 as more data points are towards positive region, followed by that is chromosome 16 which has the next largest impact. The risk assessment portrays that a Single Nucleotide Polymorphisms (SNPs) exhibiting substantial BETA values can serve as indicators for assessing an individual's susceptibility to varicose veins within the framework of polygenic risk scores. The investigative research states that the variants demonstrating considerable effect sizes may require additional exploration to uncover the underlying causal genes or biological pathways.

C) Interpretation of results:

The association between chromosomes and genomics positions contribute to the effect of varicose veins that lies in the genetic architecture of the condition. The research study states that by identifying and studying the genomics region associated with varicose veins can lead to better understanding of the biological processes that contribute to the conditions which can aid to the targeted prevention and treatment strategies.

V. Future work

A real dataset collection through a survey as well as patient analysis report has to be undertaken within the locality to identify both genetic reasons and demo graphical factors that leads to varicose veins. GWAS had a worldwide report and hence a local study could provide better understanding of the traits for VVs which could benefit the locality. Further, data analytical techniques can be applied to get better insights regarding the intensity of the factors both medically and demographically that leads to varicose veins particularly among women. Furthermore, data analytics on semi-structured and unstructured data sets could provide coherent measures to bring down the intensiveness of the varicose veins.

VI. Conclusion:

The advancement of technology that is increasing day by day brings lots of useful insights in the field of medical and healthcare that can help prevent the prevalence of various disorders. Early detection of certain disorders could layover major loss and thus can lead to the betterment of healthy life. Varicose veins symptoms thus can also be identified and witnessed at the earlier stages using data analytics especially in women as being the largest victims as compared to their counterpart.

References

[1] Patrick H. Carpentier MD , Hildegard R. Maricq MD , Christine Biro MD , Claire O. Ponçot-Makinen MD , Alain Franco MD , “Prevalence, risk factors, and clinical patterns of chronic venous disorders of lower limbs: A population-based study in France”, Journal of Vascular Surgery, Volume 40, Issue 4, October 2004, Pages 650-659, https://www.sciencedirect.com/science/article/pii/S0741521404009449

[2]Muhammad Rahil Aslam , Hafiz Muhammad Asif, Khalil Ahmad , Sana Jabbar, Abdul Hayee , Muhammad Shahid Sagheer , Jalil Ur Rehman , Sana Khalid , Abdul Sattar Hashmi , Sehrish Rana Rajpoot , Aamir Sharif , “Global impact and contributing factors in varicose vein disease development”, https://pubmed.ncbi.nlm.nih.gov/36051783/

[3] Meng-Lin Lee , Wan-Yun Chou , Yi-Wei Kao , Ming-Chih Chen , Ben-Chang Shia , Pei-Shyuan Lee , Thay-Hsiung Chen, Ting-Yuan Lin , Shih-Yu Huang , Yen-Bo Su , Wei-Che Chiu , “Varicose veins are associated with an increased risk of mitral valve regurgitation: a nationwide population-based cohort study”, https://pubmed.ncbi.nlm.nih.gov/37347156/, 2023 Aug;42(4): Pages 352-361

[4]Ann O'Banion MD, Kathleen Ozsvath MD , Bianca Cutler MSN, “A review of the current literature of ethnic, gender, and socioeconomic disparities in venous disease”, https://www.sciencedirect.com/science/article/abs/pii/S2213333X23001300

[5] Medhat Taha , et al., “Knowledge and Awareness of Varicose Veins and Its Risk Factors in Al-Qunfudah, Saudi Arabia”, https://pmc.ncbi.nlm.nih.gov/articles/PMC11170543/

[6] Waheed-Ul-Rahman Ahmed, et al., June 2022, “Genome-wide association analysis and replication in 810,625 individuals with varicose veins”, https://www.nature.com/articles/s41467-022-30765-y

[7]Wiberg, A, Ahmed, W, Furniss, D, "https://www.kaggle.com/datasets/mariusc/varicose-veins-gwas"

[8] Ikpe Justice Akpan, BSc, MSc, Ph.D. and Obianuju Genevieve Aguolu, MBBS, MPH, Ph.D., "The Application of Data Science Techniques and Algorithms in Women's Health Studies" , https://doi.org/10.1101/2022.03.27.22273006

[9] Borges do Nascimento IJ, Marcolino MS, Abdulazeem HM, Weerasekara I,Azzopardi-Muscat N, Gonçalves MA, Novillo-Ortiz D, "Impact of Big Data Analytics on People's Health: Overview of Systematic Reviews and Recommendations for Future Studies", J Med Internet Res 2021;23(4):e27275, doi: 10.2196/27275

[10] Jayasri N.P.∗ , R. Aruna, VelTech Rangarajan Dr.Sagunthala R & D Institute of Science and Technology, Chennai, Tamil Nadu, India, " Big data analytics in healthcare by data mining and classification techniques", https://www.sciencedirect.com/science/article/pii/S2405959521000849

[11] S. Lokesh , Sudeshna Chakraborty , Revathy Pulugu , Sonam Mittal , Dileep Pulugu , R. Muruganantham f , "AI-based big data analytics model for medical applications", https://www.sciencedirect.com/science/article/pii/S2665917422001684

[12] Nisrine Berros , Fatna El Mendili , Youness Filaly and Younes El Bouzekri El Idrissi, "Enhancing Digital Health Services with Big Data Analytics", https://www.mdpi.com/2504-2289/7/2/64

[13] Ping Guo, Qin Fang, Yan Wang, "Associations between varicose veins and heart failure: A genetic correlation and mendelian randomization study", https://pmc.ncbi.nlm.nih.gov/articles/PMC11098184/

[14] Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla, "Artificial Intelligence, Blockchain, Computing and Security SET", May, 2024 by CRC Press

[15] Heling Bao 1, Hui Liu 1,*, Linhong Wang, "Using Healthcare Big Data Analytics to Improve Women's Health: Benefits, Challenges, and Perspectives", Mar, 2024, https://pmc.ncbi.nlm.nih.gov/articles/PMC10960518/

[16] Mary Regina Boland et al., "Advanced Methods for Big Data Analytics in Women's Health", Pacific Symposium on Biocomputing 26, pp:50-54 (2021), https://psb.stanford.edu/psb-online/proceedings/psb21/intro-big_data.pdf