Iran Occupational Health

Original Article
http://ioh.iums.ac.ir

Iran Occupational Health.2019 (Apr-May);16(1):72-89.

Identification of the most important factors of ethnic differences in anthropometric dimensions of Iranian workers using the decision tree

Akhtar Mohamady, MSc, Department of Occupational Health, Tabriz University of Medical Sciences, Tabriz, Iran
Mahmood Reza Azghani, Associate Professor, Department of Biomechanical Engineering, Sahand University of Technology, Tabriz, Iran
Jalil Nazari, (*Corresponding author) Associate Professor, Department of Occupational Health, Tabriz University of Medical Sciences, Tabriz, Iran nazari_j@yahoo.com
Adel Mazloumi, Associate Professor, Department of Occupational Health, Teheran University of Medical Sciences, Teheran, Iran

Abstract

Background: Anthropometry is a branch of human science that considers the physical measurement of the human body, especially size and shape. One application of anthropometrical data in ergonomics is to design working space and the development of industrialized products. Therefore, anthropometrical study on tools, equipment and workstations, which are designed based on the physical dimensions of the workers, can increase productivity. According to statistics, 32% of workers possess awkward postures, and 15% use improper tools while working. These increases the possibility of work-related injuries, particularly musculoskeletal disorders. Therefore, most countries have made great efforts to build own anthropometrics databases for various groups of citizens. Due to the lack of comprehensive databases in Iran, often information from the Western countries, especially the United States are being used. The western nations anthropometric dimensions differ with that of the Iranian population, and western manufacturers design and develop tools and machines based on their own criteria. The mismatching of these designed tools and workstations with the dimensions of the Iranian user's body can cause complications such as fatigue or other physical damage. So, Iranian researchers have suggested that at least one comprehensive, up-to-date and general national database is needed. The results of several studies have shown a correlation between the dimensions of a body. These correlations can be used to create regression equations for estimating anthropometric dimensions. Identifying, categorizing, and determining the type of relationship between anthropometric dimensions play a significant role in treatment, fitness talent, and clothing production, etc. But, the relationship between these dimensions is affected by the environmental, economic and social factors' change. The important affecting physical factors include age, sex, race, body structure, occupation, diet, and physical activity. Among these factors, the race has a very critical role in the variation of body size. So, the differences among diverse races are more than the variations between different nations. Ethnic diversity is a crucial factor that can affect anthropometric data and its application areas. For example, this variety in body dimensions between people with different sex and race can produce many problems in product design. Therefore, of a modern method is essential to build a comprehensive anthropometric database from different races in the country. This is to consider the dificulties of extracting appropriate information from massive data and transforming them into knowledge on one hand, and time-consuming and high expense process for collecting data on the other hand, especially in the Iranian population with many races. Data mining is a new method used to extract useful and unknown information from raw data. Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. Decision Tree is one of the easiest and popular classification algorithms to understand and interpret. It can be utilized for both classification and regression kind of problem. It can also remove part of the consumed time and expense of collecting anthropometric data. Therefore, this study was intended to investigate and identify the most important factors of ethnic differences in anthropometric dimensions of Iranian workers using the decision tree.

Materials and Methods: The present research is a methodological study using the classification systems in the field of data mining. The sample of the study was anthropometric raw data (37 dimensions) of 3720 subjects (3,000 men and 720 women) from six races (Fars, Turks, Kurds, Lars, Baluch, and Arabs) of Iranian workers. The decision tree (DT) method was used to identify the most important factors of racial differences in anthropometric dimensions of Iranian workers. The WEKA software (version 3.6.12) was used to analyze data and implement data mining algorithms. In the case of the WEKA system, the data is extracted and translated into a standard format we call ARFF, (Attribute-Relation File Format). In the current study, the preprocessing operation on raw data performed by using a classification. Data preprocessing is one of the most critical steps in a data mining process which deals with the preparation and transformation of the initial dataset. Data preprocessing methods are divided into, data cleaning, data integration, data transformation, data reduction categories. Data cleaning routines work to clean the data by filling in missing values, smoothing noisy data, identifying or removing outliers, and resolving inconsistencies. Dirty data can cause confusion for the mining

Keywords
Anthropometry,
Data mining,
Decision tree,
Ethnic

Received: 10/12/2017
Accepted: 14/01/2019

procedure, thus removing these data is a more appropriate method. For data preparation, first of all, missing data values identified by using statistical methods and then to replacing the lost data, the PLS Filter and Missing Class Values algorithms were used in the WEKA software. These algorithms first classify the data and compute the average value for each class and fill the missing data of each class with the mean value of that class. Distorted data also eliminated by using the Remove Useless algorithm. Later data preparation, Principal Components Analysis (PCA), one of the most used methods, implemented. It is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables (entities each of which takes on various numerical values) into a set of values of linearly uncorrelated variables called principal components. The primary motivation behind PCA is to reduce, or summarize, a large number of variables into a smaller number of derived variables that may be readily visualized in 2- or 3-dimensional space. The new set of variables created by PCA can be used in other analyses, but most commonly as a new set of axes on which to plot your multivariate data. After data preparation, the data are standardized and then the KMO factor is calculated. The KMO factor measure of sampling adequacy and compares the observed correlations and partial correlations among the original variables. The KMO increases with an increase of the number of variables and the correlation coefficients between them but does not much depend on the sample size. Specifically, correlation matrices with KMO< 0.5 are entirely inappropriate whereas those with KMO below 0.6–0.7 must be treated with caution. Thereafter, the decision tree technique was applied. This technique is a tree-like structure that describes the set of rules that led to the decision, and the ease of interpretation is one of its most important features. This technique is used for categorization and is a graphical method for comparing alternative competition and assigning value to them by combining uncertainties, costs, and repayments by specific number values, which usually consists of several nodes called nodes Inputs and outputs are known. The rules created in the decision tree are also "if" and "then". In each node, you can also split more than two. The CHAID algorithm used to splits the target into two or more categories that are called the initial, or parent nodes.

Results: For the analysis of the main factors, the 37 variables (anthropometric dimensions) converted to standard values. KMO value for the total data was 0.947. Since the obtained KMO (0.95) value was greater than the considered criterion (0.7). The existence of the necessary correlation between the input variables confirms the main factors analysis. The result of the analysis indicated, of the 37 anthropometric input variables, only 21 variables had a coefficient of above 0.7 in the seven extracted factors. The remaining 16 variables had little importance due to their high correlation with the main variables and excluded from the analysis cycle. In the decision tree, the race variable selected as the target variable or dependent variable, and its relation with 21 anthropometric variables examined. The optimal number of clusters obtained was 7 clusters. After entering and determining the type of data and dividing them into two sets: one with 70 percent of the source data, for training the model, and one with 30 percent of the source data, for testing the model. This default was chosen because a 70-30 ratio is often used in data mining. The results of the study disclosed that the distinctive factors in categorizing and creating ethnic differences among men were the Shoulder height in sitting position, the eye height in sitting position, the elbow- fingers length, bi-acromial breadth, the width of foot and length of the head, and among the women, Face Breadth, bi-acromial breadth and the length of the elbow-tip of the fingers.

Conclusion: Identification of distinctive factors in the classification of racial differences is one of the main findings of the present research in various ethnic groups. According to the results of this study, these factors in Iranian women and men were different and regardless of race, the following conclusions are made; the distinguishing factors in men are related to anatomical factors (temple) and in women related to facial aesthetics (beauty). The results of this study also confirm the utility of using the decision tree method to investigate the interactions between predictor variables that can be identified the most important determinants of race by combining different nodes in the overall structure. The results from this study verified the using decision tree as an effective method of identifying important variables for classification and detection of racial differences in the anthropometric field.

Furthermore, this finding could be used to designing optimal ergonomic workstations for Iranian workers with different Races, moreover to applications in the manufactures and design process, can be used in other cases such as forensic medicine for diagnosis and making orthopedic products.

Conflicts of interest: None

Funding: Tabriz University of Medical Sciences

How to cite this article:

Mohamady A, Azghani MR, Nazari J, Mazloumi A. Identification of the most important factors of ethnic differences in anthropometric dimensions of Iranian workers using the decision tree. Iran Occupational Health.2019 (Apr-May);16(1):72-89.

*This work is published under CC BY-NC-SA 1.0 licence

مقاله پژوهشی
http://ioh.iums.ac.ir

سلامت کار ایران

دوره ۱۶، شماره ۱، فروردین و اردیبهشت ۱۳۹۸

شناسایی مهم‌ترین فاکتورهای تفاوت‌های نژادی در ابعاد آنتروپومتری کارگران ایرانی با استفاده از درخت تصمیم‌گیری

اختر محمدی: کارشناسی ارشد ارگونومی، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، تبریز، ایران

محمود رضا آذغانی: دانشیار، گروه مهندسی بیومکانیک، دانشکده مکانیک، دانشگاه سهند تبریز، تبریز، ایران

جلیل نظری: (*نویسنده مسئول) دانشیار، گروه مهندسی بهداشت حرفه‌ای، دانشکده بهداشت، دانشگاه علوم پزشکی تبریز، تبریز، ایران nazari_j@yahoo.com

عادل مظلومی: دانشیار، گروه مهندسی بهداشت حرفه‌ای، دانشکده بهداشت، دانشگاه علوم پزشکی تهران، تهران، ایران

چکیده

زمینه و هدف: آنتروپومتری به‌عنوان یکی از شاخه‌های فیزیکال آنتروپولوژی از جمله مهم‌ترین حیطه‌های ارگونومی است که به اندازه‌گیری ابعاد و ویژگی‌های بدن انسان (طول، وزن، حجم اندام‌ها، فضای حرکتی و زوایای حرکتی هر یک از این ابعاد) می‌پردازد. این داده‌ها برای تعیین شکل و اندازه ابزار و وسایل، مورد استفاده افراد در محیط کار قرار می‌گیرد، تا وسایل، ابزارآلات و ایستگاه‌کاری متناسب با ابعاد بدنی شاغلین به جهت افزایش بازده و بهره‌وری طراحی و ساخته شود. همچنین این داده‌ها می‌تواند در انتخاب افراد برای مشاغل و همسان‌کردن توانمندی هر فرد با انرژی مورد نیاز شغل محوله استفاده شود. بنابراین اندازه‌گیری دقیق ابعاد بدنی و بروز کردن داده‌ها در قالب بانک‌های اطلاعاتی تن‌سنجی (آنتروپومتری)، ضروری است. اگرچه این بانک‌های اطلاعاتی پایه و اساس طراحی ارگونومیکی محیط‌های کاری و محصولات تولیدی می‌باشند. اما انبوه داده‌های جمع‌آوری شده از ابعاد بدنی و تبدیل آنها به دانش، استفاده از روش‌های نوین را ضروری نموده است. داده‌کاوی یکی از روش‌های نوین بوده که در فضای مدیریت دانش در به دست آوردن اطلاعات نهان به کار گرفته شده و توانایی جداسازی بخش مفید اطلاعات و انجام پردازش یکسان روی داده‌ها را دارد. لذا مطالعه حاضر جهت شناسایی مهم‌ترین فاکتورهای تفاوت‌های نژادی در ابعاد آنتروپومتری کارگران ایرانی با استفاده از درخت تصمیم‌گیری طراحی و اجرا گردید.

روش بررسی: پژوهش حاضر به عنوان یک بررسی متدولوژیک در زمینه داده‌کاوی و استفاده از روش‌های دسته‌بندی (درخت تصمیم) قابل طرح می‌باشد. در پژوهش حاضر، داده‌های خام جمع‌آوری شده بدنی مشتمل بر ۳۷ بعد از ابعاد ۳۷۲۰ نفر، (۳۰۰۰ مرد و ۷۲۰ زن) در محدوده سنی ۶۰-۲۰ ساله کارگر ایرانی در برگیرنده‌ی شش قوم، فارس، ترک، کرد، لر، بلوچ و عرب گردآوری و در مطالعه‌ای با استفاده از مدل درخت تصمیم برای شناسایی مهمترین فاکتورهای تفاوتهای نژادی در ابعاد آنتروپومتری بررسی شد. جهت تحلیل داده‌ها و اجرای الگوریتم‌های داده‌کاوی نیز از نرم افزار WEKA (نسخه 3.6.12) استفاده شد.

یافته‌ها: براساس نتایج مطالعه در میان مردان، ارتفاع شانه نشسته، ارتفاع دید نشسته، طول‌آرنج-نوک انگشتان، پهنای شانه بین دو زائده آکرومیون، عرض پا و طول سر به ترتیب از عامل‌های متمایزکننده در دسته‌بندی و ایجاد تفاوت‌های نژادی در گروه‌های مورد مطالعه بودند. درحالیکه در میان زنان عامل‌های پهنای-صورت، پهنای شانه بین دو زائده آکرومیون و طول آرنج-نوک انگشتان از عامل‌های متمایزکننده در دسته‌بندی و باعث تفاوت‌های نژادی می‌باشند.

نتیجه‌گیری: نتایج مطالعه حاضر می‌تواند تاکید مجددی بر سودمندی بکارگیری روش درخت تصمیم برای بررسی تداخلات بین متغیرهای پیشگویی‌کننده باشد که می‌توان در آن با ترکیب تنوع گروها در ساختار کلی، مهمترین عوامل تعیین‌کننده نژادی را شناسایی نمود. با در نظر گرفتن هدف مطالعه حاضر و نتایج حاصل از درخت تصمیم، می‌توان نتیجه گرفت که فارغ از نژاد، عامل‌های تمایزکننده در مردان مربوط به عامل‌های تن‌سنجی (هیکل) و در زنان مربوط به عامل‌های تن-سنجی صورت (زیبایی) می‌باشند. براساس نتایج بدست آمده، که عوامل متمایز کننده در مردان و زنان جامعه ایرانی متفاوت از هم بوده، لذا این نتایج بدست آمده می‌تواند علاوه برکاربرد در صنایع تولیدی و طراحی، در موارد دیگری همچون تشخیص هویت در پزشکی قانونی، ساخت محصولات ارتوپدی نیز کاربرد داشته باشد.

کلیدواژه‌ها

آنتروپومتری،

داده‌کاوی،

درخت تصمیم،

نژاد

تاریخ دریافت: ۹۶/۰۹/۱۹

تاریخ پذیرش: ۹۷/۱۰/۲۴

تعارض منافع: گزارش نشده است.

منبع حمایت کننده: دانشگاه علوم پزشکی تبریز

شیوه استناد به این مقاله:

Mohamady A, Azghani MR, Nazari J, Mazloumi A. Identification of the most important factors of ethnic differences in anthropometric dimensions of Iranian workers using the decision tree. Iran Occupational Health.2019 (Apr-May);16(1):72-89.

***انتشار این مقاله به صورت دسترسی آزاد مطابق با CC BY-NC-SA 1.0 صورت گرفته است.**

مقدمه

آنتروپومتری به عنوان یکی از شاخه‌های فیزیکال آنتروپولوژی از جمله مهمترین حیطه‌های ارگونومی است که به اندازه‌گیری ابعاد و ویژگی‌های بدن انسان از قبیل طول، وزن و حجم اندام‌ها، فضای حرکتی و زوایای حرکتی هر یک از این ابعاد می‌پردازد. در نهایت نتایج منتج از این اندازه گیری‌ها (داده‌های آنتروپومتریکی) جهت تعیین شکل و اندازه ابزار و وسایل، مورد استفاده افراد در محیط کار قرار می‌گیرد، تا وسایل، ابزارآلات و ایستگاه کاری متناسب با ابعاد بدنی شاغلین به جهت افزایش بازده و بهره‌وری طراحی و ساخته شود(۱). همچنین داده های حاصل از اندازه‌گیری این ابعاد می‌تواند در انتخاب افراد برای مشاغل و همسان‌کردن توانمندی هر فرد با انرژی مورد نیاز شغل محوله استفاده شود (۲).

براساس آمارهای منتشره ۳۲٪ از شاغلین در هنگام کار دارای وضعیت بدنی نامناسب بوده و ۱۵٪ از آنان نیز از ابزارکاری نامناسب استفاده می‌کنند. فقدان مقیاس‌های لازم یا عدم بکارگیری داده‌های آنتروپومتری در طراحی ایستگاه‌ها و ابزارهای‌کاری متناسب با ابعاد بدنی کارگران ایرانی را می‌توان از دلایل عمده این امر به حساب آورد (۳) که این امر موجب وارد آمدن فشار وضعیتی بر بدن فرد شده و ضمن تهدید سلامت آنان، ریسک ابتلا به آسیب‌های شغلی به ویژه اختلالات اسکلتی-عضلانی را افزایش می‌دهد (۱و ۴-۵). بنابراین جهت سلامتی و راحتی نیروی انسانی شاغل در صنایع و مشاغل اداری پیشنهاد و طراحی ساخت ابزارالات (از جمله صندلی) متناسب با ابعاد کاربران ایرانی توصیه شده است (۶). به این دلیل اکثر کشورها تلاش زیادی را در جهت تدوین بانک اطلاعاتی مربوط به ابعاد آنتروپومتریکی برای گروه‌های متفاوتی از شهروندان، افراد نظامی، دانش آموزان و کارگران خود نمودند (۷). در این راستا اگر چه در کشور ایران نیز تاکنون مطالعاتی در جهت تعیین ابعاد آنتروپومتریکی در برخی جمعیت‌ها صورت گرفته که بیشتر این مطالعات نیز در گروه‌های کوچک خاصی از قبیل دانش آموزان، نظامیان، دانشجویان بوده است. با توجه به عدم وجود بانک‌اطلاعاتی جامع در این خصوص که حاوی ابعاد بدن جمعیت ایرانیان باشد، بیشتر به اطلاعات موجود مربوط به کشورهای غربی و به ویژه آمریکا مراجعه می‌شود (۷). اما به علت تفاوت‌های زیادی که ابعاد آنتروپومتریکی این جوامع با ابعاد جمعیت ایرانی دارند و همچنین شرکت‌های سازنده نیز، بیشتر طراحی ابزارآلات ساخته شده را براساس معیارهای ذهنی خود انجام می‌دهند. به دلیل عدم تناسب ابعادی این ابزارها و ایستگاه‌های کاری طراحی شده با ابعاد بدن کاربر ایرانی، عوارضی از قبیل خستگی زودرس و یا صدمات جسمانی متعدد دیگری را ایجاد می‌نمایند (۳). لذا ایجاد پایگاه‌های داده های تن‌سنجی ضروری است. بطوریکه ضرورت تهیه حداقل یک پایگاه داده جامع، به روز و غیراختصاصی از اطلاعات آنتروپومتریک جمعیت عمومی کشور را توسط محققین ذیربط پیشنهاد نمودند (۸). اما ایجاد این پایگاه‌ها (داده‌های تن‌سنجی) معمولا یک کار هزینه‌بر، وقت‌گیر، فشرده و سخت می‌باشد. لذا اگر بتوان داده‌های قدیمی تن‌سنجی را بدون فرایندهای طولانی اندازه‌گیری، و براحتی به روز کرد، بسیار ارزان‌تر خواهد شد (۹).

نتایج مطالعات (۱۰-۱۳) نشان‌دهنده وجود همبستگی بین ابعاد مختلف بدنی با یکدیگر بوده، که از این همبستگی‌ها می‌توان به منظور ایجاد معادلات رگرسیونی و به دنبال آن تخمین ابعاد آنتروپومتریکی استفاده کرد. استفاده از این روش‌ها می‌تواند بخشی از فرایند زمانبر و پرهزینه جمع‌آوری داده‌های آنتروپومتریکی را، حذف نماید.

شناسایی، طبقه‌بندی و مشخص نمودن نوع ارتباط بین ابعاد آنتروپومتریک، می‌تواند اثر قابل توجهی را در درمان، استعداد یابی ورزشی، تولید پوشاک و ... بازی کند، ولیکن ارتباطات بین این ابعاد (آنتروپومتریک) تحت تاثیر عوامل محیطی، اقتصادی و اجتماعی در حال تغییر می‌باشد (۱۴).

تفاوت گروه‌ها در ابعاد بدنی از عامل‌های مختلف بیولوژیکی و محیطی متاثر است. این تفاوت‌ها می‌تواند تحت تاثیر شرایط محیط جغرافیایی و همچنین شرایط اجتماعی باشد. به طور مثال گروه‌های قومی که در آب و هوای گرم زندگی می‌کنند دارای نسبت سطح بدن به جرم بدن بیشتری هستند. براین اساس نتایج مطالعات پیشین پژوهشگران نشان داده که داده‌های آنتروپومتریکی در جمعیت‌های موجود در مناطق

مختلف دارای اختلاف وسیعی می‌باشند. بعنوان مثال جمعیت اسکاندیناویایی دارای قد بلندتری هستند، اما آسیایی‌ها و ایتالیایی‌ها قد کوتاه‌تری دارند (۱۵). در میان ملت‌های مختلفی نیز که دارای نژاد یکسانی هستند، ممکن است به خاطر تفاوت در شرایط اقتصادی و اجتماعی، ابعاد و نسبت های بدنی متفاوتی وجود داشته باشد (۱۶). به طورکلی فاکتورهای موثر بر روی اندازه‌های بدنی شامل سن، جنس، نژاد، ساختمان بدن، حرفه، رژیم غذایی و فعالیت‌های فیزیکی بوده که از بین این عوامل، نژاد نقش بسیار مهمی در اختلاف ابعاد بدن افراد دارد (۱۵). لذا تفاوت‌های آنتروپومتریکی در میان نژادهای گوناگون بیشتر از تفاوت‌های موجود میان ملل مختلف می‌باشد (۱۶).

اصطلاح قومیت به گروهی اشاره دارد که با ویژگی‌های بنیادی از قبیل زبان، آداب و رسوم و میراث تاریخی، از سایر گروه‌های اجتماعی که دارای پیوستگی و همبستگی نژادی هستند، متمایز می‌باشند. ایران از جمله کشورهایی است که از تنوع قومی در ساختار جمعیتی خود برخوردار بوده و اقوام مختلف اعم از فارس، آذری، کرد، لر، بلوچ، ترکمن و عرب به شکل اقلیت‌های مشخص در مناطق و استانهای مختلف پراکنده شده‌اند (۱۷).

تنوع قومیتی فاکتور مهمی است که می‌تواند در داده-های آنتروپومتریکی و حوزه‌های کاربرد آن تاثیرگذار باشد. بعنوان مثال این تغییرپذیری در ابعاد بدنی در بین افراد با جنسیت‌های مختلف و در بین نژادهای مختلف می‌تواند طراحی محصول را با مشکلاتی روبرو کند (۱۸). بنابراین با توجه به ضرورت داشتن بانک اطلاعاتی جامع مربوط به ابعاد آنتروپومتریکی نژادهای مختلف در کشور و استخراج اطلاعات مناسب از میان انبوه داده‌ها و تبدیل آنها به دانش از سویی، و فرایند زمانبر و پرهزینه بودن جمع‌آوری داده‌های تن‌سنجی به خصوص در جمعیت ایرانی با تنوع نژادی بالا از سوی دیگر، استفاده از روش‌های نوین ضروری می‌باشد.

در دسترس بودن حجم بزرگی از داده‌های خام حاصل از اندازه‌گیری ابعاد و ویژگی‌های بدن انسان در نژادهای مختلف، اگر چه فرصت و مجال است، ولیکن مجموعه عظیم این داده‌ها به خودی خود کاربردی نداشته و دانشی که در آن وجود دارد، بطور معمول نهان و استفاده نشده است. برای معنی بخشیدن به این داده‌ها باید آنها را تحلیل و تبدیل به اطلاعات و یا بهتر از آن، یعنی تبدیل به دانش کرد (۱۹). بنابراین استخراج اطلاعات مناسب از میان انبوه داده‌ها و تبدیل آنها به دانش، نیازمند استفاده از روش‌های نوین می‌باشد. داده‌کاوی یکی از این ابزارها و رویکردها و بعبارت دیگر از روش های نوین بوده که در فضای مدیریت دانش در به دست آوردن این اطلاعات نهان به کار گرفته می‌شود (۲۰). بعبارت دیگر به دنبال استخراج اطلاعات مفید و ناشناخته از داده های خام می‌باشد (۲۱-۲۲). داده‌کاوی که به بررسی و تجزیه تحلیل تعداد زیادی از داده‌ها برای کشف الگوها و قوانین معنی‌دار اطلاق می‌شود، تشخیص الگو را به صورت خود به خود انجام داده و تلاش می‌کند الگوهایی را در داده‌ها کشف نماید که با روش‌های آماری مرسوم به سختی قابل‌کشف هستند.

داده‌کاوی توانایی جداسازی بخش مفید اطلاعات و انجام پردازش یکسان روی داده‌ها را داشته و دسته-بندی یکی از عملکردهای پیشگویانه آن می‌باشد. دسته‌بندی به عمل تقسیم جمعیت ناهمگن به تعدادی از زیرمجموعه‌ها یا گروه‌های همگن گفته می‌شود، و در واقع راهی را برای یافتن ساختار داده‌های پیچیده فراهم می‌کند. دسته‌بندی فرایند، یافتن مدلی است که با تشخیص دسته‌ها و یا مفاهیم‌داده، می‌تواند دسته ناشناخته اشیا دیگر را پیشگویی نماید (۲۳-۲۴). دسته‌بندی به عنوان قالب مفهومی و الگوریتمی غنی برای تحلیل و تفسیر داده‌ها مطرح شده است. دسته‌بندی شکل‌های مختلف اندام انسان، با کمک آنالیز داده‌های مربوط به قسمت‌های مختلف بدن انجام می‌گیرد. بکارگیری تکنیک‌های داده کاوی روی این داده‌ها و استخراج قواعد معنی‌دار آنها، می‌تواند جهت دسته‌بندی شکل‌های مختلف، مفید واقع گردد (۲۷-۲۵). یکی از روش‌های رایج دسته‌بندی، درخت تصمیم بوده که به یک انتخاب خوب کمک می‌کند، به خصوص تصمیم‌گیری‌هایی که در بردارنده هزینه بالا و خطرات زیادی می‌باشند. لذا هدف مطالعه حاضر بر این است که با استفاده از مدل درخت تصمیم‌گیری، ارتباط‌های مهم در تفاوت‌های نژادی در میان قومیت‌های ایرانی را بررسی و شناسایی نمایند. انتظار می‌رود دانش بدست آمده از نتایج این مطالعه بتواند روش موثری را برای شناسایی متغییرهای مهم در دسته

بندی نوع بدن را فراهم نماید.

روش بررسی

پژوهش حاضر به عنوان یک بررسی متدولوژیک در زمینه داده‌کاوی و استفاده از روش‌های دسته‌بندی (درخت تصمیم) قابل طرح می‌باشد. با استفاده از داده‌های خام جمع آوری شده مربوط به ۳۷۲۰ نفر افراد نمونه، شامل ۳۰۰۰ مرد و ۷۲۰ زن در محدوده سنی ۶۰-۲۰ ساله کارگر ایرانی در برگیرنده‌ی شش قوم، فارس، ترک، کرد، لر، بلوچ و عرب گردآوری شده، مطالعه‌ای با استفاده از مدل درخت تصمیم برای شناسایی مهمترین فاکتورهای تفاوتهای نژادی در ابعاد آنتروپومتری کارگران ایرانی طراحی شد. که جزئیات دقیق روش تعیین حجم نمونه از قومیت های مختلف، نحوه نمونه‌گیری، ابعاد، وسایل و تجهیزات بکار رفته برای اندازه‌گیری ۳۷ بعد مورد بررسی در مقاله چاپ شده (۲۸) گزارش شده است.

جهت تحلیل داده‌ها و اجرای الگوریتم‌های داده‌کاوی از نرم افزار WEKA (نسخه 3.6.12) استفاده شد. این نرم‌افزار الگوریتم‌های معروف زیادی را برای طبقه‌بندی، خوشه‌بندی، استخراج قوانین انجمنی و .. به صورت آماده مهیای استفاده دارد. این نرم افزار در دانشگاه وایکاتو در نیوزیلند پیاده سازی شده و قالب دریافت اطلاعات در آن نیز ARFF است.

جهت کشف دانش براساس داده‌ها با استفاده از دسته‌بندی، ابتدا بایستی عملیات پیش‌پردازش روی داده‌های خام انجام شود. پیش پردازش برای داده‌های خام شامل پاکسازی داده‌ها و تبدیل داده‌ها می‌باشد. پاکسازی داده‌ها، شامل جایگزینی داده‌های گمشده و حذف داده‌های آلوده به خطا است. که داده‌های گمشده را می‌توان با قرار دادن مقادیر شایع پراحتمال(مد) جایگزین نمود. در داده‌های آلوده به خطا که در آن یک تغییر ناگهانی وجود دارد. این خطا با استفاده از روش تحلیل داده‌های پرت رفع می‌شود. که در این روش می‌توان داده آلوده به خطا را حذف، یا با جایگزینی مقادیر ممکن اصلاح نمود، اما حذف نمودن این داده‌ها، روش مناسب‌تری است.

در مطالعه حاضر برای آماده‌سازی داده‌ها، ابتدا با استفاده از روش‌های آماری، مقادیر داده‌های مفقوده شناسایی و برای جایگزین نمودن داده‌های گمشده نیز از الگوریتم‌های PLSFilter و Missing class values در نرم‌افزار WEKA استفاده گردید. این الگوریتم‌ها ابتدا داده‌ها را طبقه‌بندی نموده و برای هر طبقه مقدار میانگینی را محاسبه نموده و داده مفقوده هر کلاس را با مقدار میانگین آن کلاس پر نمود. داده‌های پرت و مغشوش نیز توسط الگوریتم Remove Useless هموار شدند. بعد از آماده‌سازی داده‌ها، تجزیه و تحلیل عامل‌های اصلی (PCA) انجام گرفت. در این روش متغیرهای موجود در یک فضای چند حالته همبسته به یک مجموعه از مولفه‌های غیرهمبسته خلاصه شدند که هریک از آنها ترکیب خطی از متغیرهای اصلی می‌باشند. این مولفه‌های غیرهمبسته بدست آمده مولفه‌های اصلی نامیده می‌شوند که از بردارهای ویژه ماتریس کواریانس یا ماتریس همبستگی متغیرهای اصلی حاصل می‌شوند. از کاربردهای مهم این روش می‌توان به کاهش ابعادی متغیرها و یافتن ساختار ارتباطی بین آنها اشاره نمود. در این روش محورهای مختصات جدیدی برای داده‌ها تعریف شده و داده‌ها براساس این محورهای مختصات جدید بیان می‌شوند. اولین محور باید در جهتی قرار گیرد که واریانس داده‌ها بیشینه شود (یعنی در جهتی که پراکندگی داده‌ها بیشتر است). دومین محور باید عمود بر محور اول به گونه‌ای قرار گیرد که واریانس داده‌ها بیشینه شود. به همین ترتیب محورهای بعدی عمود بر تمامی محورهای قبلی به گونه‌ای قرار می‌گیرند که داده‌ها در آن جهت، دارای بیشترین پراکندگی باشند. مراحل تجزیه و تحلیل عاملی انجام شده در شکل ۱ ارائه شده است. در مرحله نخست داده‌ها استانداردسازی شده و سپس عامل KMO محاسبه می‌گردد. در صورتی که عامل KMO کمتر از ۰/۵ باشد، داده‌ها برای تجزیه و تحلیل عوامل اصلی مناسب نخواهد بود ولیکن اگر مقدار آن بزرگ‌تر از ۰/۷ باشد همبستگی‌های موجود در بین داده‌ها برای تجزیه و تحلیل مناسب خواهد بود.

با استفاده از این روش، ترکیباتی از P متغیر اولیه $X_1,X_2,.....X_p$ برای ایجاد حداکثر P عامل مستقل به صورت $PC_1, PC_2, ..., PC_p$ ایجاد می‌شود. هر عامل اصلی می‌تواند به صورت دنباله‌ای که در رابطه ۱ نشان داده شده، مشخص گردد.

شکل ۱- مراحل تحلیل‌مولفه‌های‌اصلی (PCA)

متغیرهای $PC_1, PC_2, \ldots PC_p$ را عامل‌های اصلی گویند. ضرایب a_{ij} طوری تخمین‌زده می‌شوند که برای اولین عامل حداکثر واریانس داده‌ها در نظر گرفته (اولین عامل اصلی حداقل با تعدادی از متغیرها همبسته می‌باشد) و دومین عامل، بیشترین واریانسی که توسط عامل اصلی اول محاسبه نشده را در نظر می‌گیرد (دومین عامل اصلی با تعدادی از متغیرهای مشاهده شده که همبستگی بالایی با مولفه اول را ندارند، همبسته است و همچنین عامل دوم با عامل اول همبستگی ندارند) و این روند ادامه می‌یابد تا آخرین عامل تمامی واریانس موجود در داده‌ها را در بر می‌گیرد.

$$PC_2 = a_{12}X_1 + a_{22}X_2 + \ldots + a_{P2}X_P \quad (۱)$$

از آنجائیکه تعداد عامل‌های اصلی براساس میزان پوشش پراکندگی‌ها توسط عامل‌ها مشخص می‌گردد (افزایش تعداد عامل‌ها تا جایی صورت می‌گیرد که میزان قابل توجهی از پراکندگی داده‌ها در نظر گرفته شوند) با اجرای چرخش مناسب بر روی ماتریس ضرایب عامل‌ها که به مرحله تجزیه و تحلیل عامل‌های اصلی نیز مشهور است، متغیرهایی که ضرایب بالایی در عامل‌های اصلی استخراج شده را داشته باشند را به عنوان متغیرهای مهم جهت ورود به مدل سازی انتخاب می‌گردند.

درخت تصمیم (Decision tree): این تکنیک، ساختاری شبیه به درخت است که مجموعه قواعدی که منجر به تصمیم گیری شده است را توصیف می‌کند و سهولت تفسیر از ویژگیهای مهم آن است (۲۹). از این تکنیک برای دسته‌بندی استفاده می‌شود و یک روش گرافیکی برای مقایسه رقابت جایگزین و اختصاص ارزش به آنها از طریق ترکیب عدم قطعیت‌ها، هزینه‌ها و بازپرداخت‌ها به وسیله ارزش‌های خاص عدد می‌باشد که معمولا از چندین گره تشکیل می‌شود که با نام گره-های ورودی و خروجی شناخته می‌شوند. قوانین ایجاد شده در درخت تصمیم نیز به صورت "اگر" و" آنگاه بیان می‌شوند. در هر گره، نیز می‌توان بیش از دو تقسیم داشت. یکی از الگوریتم‌های درخت تصمیم، الگوریتم CHAID[1] می باشد که این الگوریتم توسط کاس در سال ۱۹۸۰ برای استفاده در مورد متغیرهای کیفی (متغیرها یا داده‌های کیفی، داده‌هایی هستند که ارزش پاسخ‌های داده شده در آن یکسان نمی‌باشد) معرفی شده است. اما این الگوریتم می‌تواند برای متغیرهای کمی (متغیرها یا داده‌های کمی، داده‌هایی هستند که ارزش پاسخ‌های داده شده در آن یکسان نمی‌باشد) گروه‌بندی شده نیز استفاده شود.

این الگوریتم برای ساخت یک درخت تصمیم گیری، داده ها را متناوباً به زیر مجموعه‌های مشابه افراز می‌کند تا آنجا که هر زیر مجموعه دارای تعداد مشخصی نمونه شود. این الگوریتم می‌تواند درختی تولید کند که در برخی از مواقع به صورت غیر دو دویی عمل کند، در واقع از روش جدا کردن چند تایی به جای جدا کردن دو دویی استفاده می‌کند. در این روش از مقدار P-Value آماره کای-دو مربوط به آزمون استقلال جداول توافقی، استفاده می‌شود. از بین متغیرهای موجود، متغیری که دارای P-Value کوچک‌تری باشد در مرحله اول برای تقسیم‌ها روی یک گره در نظر گرفته می‌شود. به بیان دیگر، CHAID ابتدا تفاوت های هر نمونه را با سایر نمونه ها می یابد و درخت مورد نظر را تولید می کند. هرس کردن درخت از طریق یافت تفاوت های مشابه انجام می شود. جهت اعتبارسنجی مدل درختی، داده‌ها به دو بخش داده‌های آموزش و آزمون تقسیم شدند مدل درختی با استفاده از داده‌های آموزش ساخته شد و مدل ساخته شده بر روی داده‌های آزمون مورد تست قرار گرفتند. درصد نمونه‌هایی از داده‌های آزمون که ویژگی هدف آنها توسط مدل، درست تشخیص داده شده بود دقت مدل

[1] Chi-squared Automatic Interaction Detector

را بیان می‌کند (۳۰).

یافته‌ها

در این بررسی، داده های مربوط به ۳۷۲۰ نفر ،۳۰۰۰ نفر مرد و ۷۲۰ نفر زن که در زمینه ابعاد آنترپومتری کارگران شاغل در محدوده سنی ۶۰-۲۰ سال گردآوری شده بود، از نظر ارتباط بین متغیرها و طراحی درخت تصمیم، استفاده شد. که جدول شماره ۱ میانگین و انحراف‌معیار جمعیت مورد مطالعه را نشان می‌دهد.

در مرحله اول بعد از آماده سازی داده‌ها، برای تجزیه و تحلیل عامل‌های اصلی ابتدا ۳۷ متغیر اولیه آنتروپومتری به مقادیر استاندارد تبدیل شد. مقدار KMO برای کل داده‌ها، برابر ۰/۹۴۷ بدست آمد. با توجه به اینکه مقدار KMO بدست آمده بزرگ‌تر از ۰/۷ بوده و وجود همبستگی لازم بین متغیرهای ورودی برای انجام تحلیل عامل‌های اصلی را تایید می‌کند.

جدول ۱- شاخص‌های آماری داده‌های آنتروپومتریکی کارگران ایرانی۲۰-۶۰ سال

متغییر آنتروپومتریکی	مرد		زن	
	میانگین	انحراف معیار	میانگین	انحراف معیار
طول قد	۱۷۱/۸۶	۷/۷۴	۱۵۸/۴۷	۶/۳۲
ارتفاع چشم	۱۶۰/۸۸	۷/۹۶	۱۴۷/۱۸	۸/۰۲
ارتفاع شانه	۱۴۳/۶۲	۷/۴۶	۱۳۰/۹۳	۵/۷۳
ارتفاع آرنج	۱۰۸/۳۱	۶/۱۸	۹۸/۹۱	۴/۹۷
ارتفاع کفل	۸۸/۰	۵/۰	۸۶/۲۴	۵/۶۹
ارتفاع برآمدگی انگشتان	۷۴/۸	۴/۶۴	۶۹/۴۸	۳/۹۸
ارتفاع نوک انگشتان میانه	۶۴/۹	۴/۵۳	۵۹/۶۱	۳/۷۴
ارتفاع نشسته	۹۱/۰	۵/۳	۸۲/۲۹	۴/۰۸
ارتفاع چشم نشسته	۸۰/۰	۴/۹۴	۷۲/۲۸	۴/۳۸
ارتفاع آرنج نشسته	۶۱/۷	۴/۴۲	۵۵/۹	۴/۳۵
ارتفاع شانه نشسته	۲۶/۹	۴/۲۸	۲۱/۵	۳/۳۴
ضخامت ران	۱۵/۲	۲/۱۶	۱۴/۷۴	۲/۴۵
طول کفل زانو	۵۷/۹۹	۳/۱۷	۵۶/۰	۳/۴۳
طول کفل رکبی	۴۵/۹۴	۳/۲۸	۴۴/۶۷	۳/۳۶
ارتفاع زانو	۵۲/۱۷	۴/۲۲	۴۷/۶۲	۳/۴۵
پهنای شانه(بین دو عضله دلتوئید)	۴۵/۱۲	۳/۳۶	۴۲/۷۵	۴/۱۱
پهنای شانه بین دو زائده آکرومیون	۳۷/۸۲	۴/۵۴	۳۵/۳۵	۳/۹
پهنای باسن	۳۸/۰	۳/۳۷	۳۶/۹۵	۳/۸۸
عمق سینه	۲۳/۵	۳/۰۶	۲۵/۵۴	۳/۹۵
عمق شکم	۲۵/۰	۳/۶۸	۲۷/۷۹	۳/۹۲
طول شانه-آرنج	۳۵/۴۳	۲/۹۶	۳۱/۸۵	۲/۷
طول آرنج-نوک انگشتان	۴۶/۱۵	۴/۲۲	۴۲/۵۴	۲/۴۹
طول اندام فوقانی	۷۷/۳۵	۴/۸۴	۷۱/۱	۴/۱۸
طول شانه-چنگش	۶۶/۴۲	۴/۸۲	۶۱/۲۴	۴/۳۸
طول سر	۱۸/۷۶	۱/۱۵	۱۸/۱۴	۱/۱۹
پهنای سر	۱۴/۷۵	۰/۹۹	۱۳/۹۳	۴/۶۱
طول دست	۱۸/۷۹	۱/۱۵	۱۷/۲۴	۰/۹۷
پهنای دست	۸/۵	۰/۶۳	۷/۵	۰/۵۳
طول کف پا	۲۵/۴۲	۱/۶۹	۲۲/۹	۱/۴۳
پهنای کف پا	۹/۶۴	۰/۷۷	۸/۶	۰/۷۵
فاصله بین نوک انگشتان میانی دست راست و چپ زمانی که بازوها کاملا باز باشند	۱۷۴/۱۵	۹/۵۵	۱۵۸/۳	۸/۷۶
فاصله بین آرنج ست راست و چپ هنگامی که بازوها به طرفین باز شده و ساعدها خم شده	۸۸/۹۵	۵/۶۶	۸۲/۸۷	۴/۹
حد دسترسی چنگش ایستاده	۲۰۸/۸	۱۰/۸۹	۱۸۹/۰	۷/۹
حد دسترسی چنگش نشسته	۱۲۷/۲۴	۷/۸۶	۱۱۳/۳۷	۶/۷۵
حد دسترسی چنگش به جلو	۷۸/۴۹	۸/۴۱	۶۸/۴۷	۵/۰
وزن	۷۴/۲۶	۱۲/۵۵	۵۹/۶۶	۱۰/۳۷

تجزیه و تحلیل عامل‌های اصلی بر روی کل داده‌ها اجرا گردید. در ادامه ماتریس کواریانس متغیرها محاسبه و بردارهای ویژه و مقادیر ویژه مربوطه محاسبه شدند. که بردارهای ویژه برای تشکیل هر عامل در جدول ۲ ارائه شده است. برای تشکیل عامل اول (PC_1) ضریب ۰/۹۴ را در طول قد و ضریب ۰/۹۲۵ را در ارتفاع چشم و ضریب ۰/۹۳۷ در ارتفاع شانه و ... اعمال گردد. به همین ترتیب برای همه متغیرهای آنتروپومتری ضرایبی محاسبه شد که ضرایب بالاتر نشان‌دهنده تاثیر بیشتر این متغیرها در تشکیل عامل اول می‌باشند.

با توجه به اینکه در روش تجزیه و تحلیل عامل‌های اصلی، متغیرهای اصلی متغیرهایی هستند که حداقل یکی از ضرایب آنها که برای تشکیل عامل مربوطه استفاده می‌شود، دارای مقادیر نسبتا بالایی باشند. با توجه به معیار درنظر گرفته شده (۰/۶۵)، این معیار براساس نتایج مطالعات پیشین انتخاب شده است(۲۳)، از ۳۷ متغیر آنتروپومتریکی ورودی، تنها ۲۱ متغیر ضریب بالای ۰/۶۵ را در هفت عامل استخراج شده، دارا بودند. که نتایج محاسبات مربوط به ماتریس دوران یافته در جدول شماره ۳ نشان داده شده است. ۱۶ متغیر باقیمانده دیگر نیز به دلیل همبستگی زیاد با متغیرهای اصلی، کم اهمیت تشخیص داده شده و از چرخه تجزیه و تحلیل خارج شدند. در مراحل اجرای درخت تصمیم نیز متغیر نژاد به عنوان متغیر هدف یا

جدول ۲- مقادیر مربوط به بردارها ویژه برای تشکیل هر مولفه

ابعاد آنتروپومتری	مولفه اول (PC_1)	مولفه دوم (PC_2)	مولفه سوم (PC_3)	مولفه چهارم (PC_4)	مولفه پنجم (PC_5)	مولفه ششم (PC_6)	مولفه هفتم (PC_7)
طول قد	۰/۹۴۰	۰/۱۵۳-	۰/۰۸۳-	۰/۰۹۱	۰/۰۷۶	۰/۰۵۳-	۰/۰۷۳-
ارتفاع چشم	۰/۹۲۵	۰/۱۴۷-	۰/۱۲۳-	۰/۱۲۵	۰/۰۴۳	۰/۰۶۲-	۰/۰۶۰-.
ارتفاع شانه	۰/۹۳۷	۰/۱۳۹-	۰/۰۶۹-	۰/۰۸۰	۰/۰۹۴	۰/۰۵۲-.	۰/۰۳۴-
ارتفاع آرنج	۰/۸۶۴	۰/۱۶۶-	۰/۲۱۸-	۰/۰۷۲	۰/۱۱۹	۰/۱۱۳-	۰/۰۴۰
ارتفاع کفل	۰/۵۷۵	۰/۱۰۴-	۰/۱۳۲	۰/۳۷۰	۰/۲۳۴	۰/۲۸۴-	۰/۰۹۵-
ارتفاع برآمدگی انگشتان	۰/۷۷۷	۰/۰۴۳-	۰/۲۶۶-	۰/۲۶۶	۰/۱۰۸	۰/۱۹۳-	۰/۰۵۵-
ارتفاع نوک انگشتان میانه	۰/۶۹۱	۰/۰۸۳-	۰/۳۲۲-	۰/۲۹۲	۰/۱۴۲	۰/۲۳۲-	۰/۰۲۴
ارتفاع نشسته	۰/۷۹۸	۰/۰۸۳-	۰/۳۱۸-	۰/۱۳۱-	۰/۱۲۸-	۰/۰۲۱	۰/۱۳۱
ارتفاع چشم نشسته	۰/۷۷۱	۰/۰۴۳-	۰/۴۰۳-	۰/۱۲۲-	۰/۱۳۵-	۰/۰۲۶-	۰/۱۰۹
ارتفاع آرنج نشسته	۰/۷۵۰	۰/۰۳۷	۰/۲۴۶-	۰/۲۸۸-	۰/۰۰۸-	۰/۱۱۱-	۰/۲۲۸
ارتفاع شانه نشسته	/۴۲۱.	۰/۱۷۱	۰/۵۷۹-	۰/۴۴۲-	۰/۰۶۰	۰/۱۵۲-	۰/۱۲۲
ضخامت ران	۰/۲۶۸	۰/۵۳۷	۰/۰۹۱	۰/۲۲۳	۰/۱۶۹-	۰/۰۰۵-	۰/۳۸۲
طول کفل زانو	۰/۵۹۹	۰/۱۳۱	۰/۲۶۴	۰/۴۰۸	۰/۰۶۹	۰/۱۶۲-	۰/۰۳۰-
طول کفل رکبی	۰/۴۱۹	۰/۲۰۳	۰/۴۱۲	۰/۱۱۴	۰/۳۳۶	۰/۴۴۴-	۰/۰۷۸-
ارتفاع زانو	۰/۷۷۷	۰/۱۱۴-	۰/۰۸۸-	۰/۱۸۷	۰/۲۴۱-	۰/۰۴۹	۰/۰۰۷-
پهنای شانه(بین دو عضله دلتوئید)	۰/۴۷۵	۰/۶۲۳	۰/۱۰۷-	۰/۰۷۰	۰/۰۳۶-	۰/۲۶۴	۰/۱۰۶-
پهنای شانه بین دو زائده آکرومیون	۰/۲۷۴	۰/۲۰۴	۰/۳۸۸-	۰/۳۱۹	۰/۲۰۰	۰/۵۶۹	۰/۲۸۰-
پهنای باسن	۰/۱۵۶	۰/۶۱۹	۰/۰۲۵	۰/۰۴۵	۰/۳۹۶	۰/۲۶۶	۰/۰۵۷-
عمق سینه	۰/۰۰۲	۰/۷۴۹	۰/۱۲۳	۰/۱۸۶	۰/۰۹۸-	۰/۰۴۴-	۰/۱۹۱
عمق شکم	۰/۱۷۷	۰/۷۸۱	۰/۰۸۳	۰/۰۴۵	۰/۰۴۰	۰/۰۳۷	۰/۲۰۴
طول شانه–آرنج	۰/۶۴۷	۰/۲۴۲-	۰/۰۷۲	۰/۰۸۷	۰/۰۷۴-	۰/۲۹۹	۰/۱۳۹-
طول آرنج–نوک انگشتان	۰/۵۸۵	۰/۱۱۱-	۰/۴۴۶	۰/۱۸۰-	۰/۲۸۱	۰/۴۷۰	۰/۲۴۶
طول اندام فوقانی	۰/۷۲۳	۰/۳۰۳-	۰/۳۰۸	۰/۱۱۲-	۰/۰۶۳	۰/۱۹۷	۰/۱۲۶
طول شانه–چنگش	۰/۶۵۱	۰/۳۲۸-	۰/۳۳۱	۰/۱۳۲-	۰/۰۳۲-	۰/۱۶۸	۰/۲۸۷
طول سر	۰/۲۳۳	۰/۴۶۲	۰/۱۰۴	۰/۳۰۱-	۰/۱۷۹	۰/۱۷۲-	۰/۴۲۴-
پهنای سر	۰/۴۴۱	۰/۳۱۸	۰/۰۲۸	۰/۴۵۸-	۰/۰۶۰-	۰/۰۲۸	۰/۰۶۴-
طول دست	۰/۶۷۳	۰/۰۵۸.	۰/۲۲۶	۰/۱۰۶-	۰/۱۸۹-	۰/۰۲۸-	۰/۳۴۶-
پهنای دست	۰/۵۳۹	۰/۳۴۰	۰/۰۲۷	۰/۰۷۰-	۰/۳۹۹-	۰/۰۴۰	۰/۳۵۳-
طول کف پا	۰/۷۰۸	۰/۰۴۲-	۰/۲۲۴	۰/۰۰۴-	۰/۳۶۷-	۰/۰۱۰-	۰/۰۸۵-
پهنای کف پا	۰/۴۴۴	۰/۳۶۰	۰/۳۰۶	۰/۰۹۹-	۰/۴۷۳-	۰/۲۳۳-	۰/۰۴۴-

ادامه **جدول ۲**

ابعاد آنتروپومتری	مولفه اول (PC_1)	مولفه دوم (PC_2)	مولفه سوم (PC_3)	مولفه چهارم (PC_4)	مولفه پنجم (PC_5)	مولفه ششم (PC_6)	مولفه هفتم (PC_7)
طول قد	۰/۹۴۰	۰/۱۵۳-	۰/۰۸۳-	۰/۰۹۱	۰/۰۷۶	۰/۰۵۳-	۰/۰۷۳-
فاصله بین نوک انگشتان میانی دست راست و چپ زمانی که بازوها کاملا باز باشند	۰/۸۴۱	۰/۱۹۱-	۰/۱۴۴	۰/۰۵۳	۰/۰۹۸-	۰/۱۲۶	۰/۰۰۵
فاصله بین آرنج ست راست و چپ هنگامی که بازوها به طرفین باز و ساعدها خم شده	۰/۶۲۷	۰/۳۲۵-	۰/۲۵۸	۰/۱۶۸	۰/۱۹۱-	۰/۳۰۱	۰/۰۶۳
حد دسترسی چنگش ایستاده	۰/۸۹۹	۰/۰۹۸-	۰/۰۳۶	۰/۱۳۹-	۰/۰۸۶	۰/۰۰۹	۰/۰۳۶-
حد دسترسی چنگش نشسته	۰/۷۵۶	۰/۱۳۰	۰/۰۳۲-	۰/۴۰۲-	۰/۰۹۴	۰/۰۶۶-	۰/۰۱۹-
حد دسترسی چنگش به جلو	۰/۵۴۰	۰/۰۶۱	۰/۲۸۶	۰/۳۴۱-	۰/۴۵۴	۰/۲۶۴	۰/۰۰۲-
وزن	۰/۵۲۵	۰/۵۱۶	۰/۰۹۳-	۰/۲۱۶	۰/۰۴۴-	۰/۱۱۴	۰/۲۷۸

متغیر وابسته انتخاب و رابطه آن با ۲۱ متغیر آنتروپومتری شامل، ارتفاع شانه، ارتفاع کفل، ارتفاع برآمدگی انگشتان، ارتفاع نوک از انگشت میانه، ارتفاع نشسته، ارتفاع چشم نشسته، ارتفاع آرنج نشسته، ارتفاع شانه نشسته، ضخامت ران، پهنای شانه بین دو زائده آکرومیون، عمق سینه، عمق شکم، طول آرنج-نوک انگشتان، طول اندام فوقانی، طول شانه-چنگش، طول سر، پهنای دست، پهنای کف پا، فاصله بین آرنج دست راست و چپ هنگامی که بازوها به طرفین باز شده و ساعدها خم شده و نژاد بررسی گردید. طبق نتایج بدست آمده از این تحقیق تعداد بهینه خوشه‌ها ۷ خوشه می‌باشد. به عبارت دیگر زمانی که داده‌ها در۷ خوشه گروه بندی شده باشند حالتی است که فاصله بین داده‌ها در یک خوشه حداقل و فاصله بین داده‌ها در خوشه‌های مختلف حداکثر می‌شود. به اختصار تحلیلی از نتایج خوشه‌بندی در ادامه داده شده است. برای استفاده از تکنیک درخت تصمیم از الگوریتم CHAID، برای مجموعه داده‌ها استفاده گردید. بعد از ورود و تعیین نوع داده‌ها و تقسیم بندی آنها به دو بخش آموزش و آزمون، این الگوریتم بر روی ۷۰٪ داده‌ها (داده‌های آموزش) اجرا شده و مدل حاصل از آنها بر روی۳۰٪ مابقی داده‌ها (داده‌های آزمون) آزمایش شده است. درخت حاصل از الگوریتم درخت تصمیم برای زنان در شکل ۲ و برای مردان در شکل ۳ نشان داده شده است. و تحلیل نتایج (قوانین) ایجاد شده توسط این الگوریتم درخت تصمیم برای زنان و مردان به اختصار و جداگانه در جداول ۴ و ۵ ارائه شده است.

براساس نتایج بدست آمده از الگوریتم درخت تصمیم در زنان شش نژاد مورد مطالعه (شکل۲)، در کل بین ۶۳۴ نفر روابط برقرار شد، که از این تعداد ۲۹۳ نفر (۴۲٪/۲) به نژاد فارس، ۲۰۰ نفر (۳۱٪/۵) به نژاد ترک، ۵۰ نفر (۷٪/۹) به نژاد عرب، ۳۴ نفر (۵٪/۴) به نژاد بلوچ، ۳۰ نفر (۴٪/۷) به نژاد کرد و ۲۷ نفر (۴٪/۳) به نژاد لر تعلق داشتند. این درخت تصمیم، با شش گره روابطی را بین زنان مورد مطالعه برقرار کرده است:

گره اول از پهنای‌صورت شروع شده، بطوری‌که ۲۳۲ نفر (۳۶٪/۶) از کل افراد، دارای پهنای‌صورت کوچکتر و یا مساوی ۱۳/۰۵ سانتی‌متر بوده که از این تعداد ۵۶٪ به نژاد ترک و ۲۱/۶٪ به نژاد فارس تعلق داشتند.

گره دوم مربوط به افرادی با پهنای صورت بزرگتر از ۱۳/۵ سانتی‌متر بوده که این زیر گروه ۴۰۲ نفر بوده که ۶۰٪/۴ از آنها متعلق به نژاد فارس و ۱۷/۴٪ به نژاد ترک بودند.

گره سوم، زیر گروهی از گره دوم با ۱۴۲ نفر که دارای پهنای‌شانه بین دو زائده آکرومیون کوچکتر یا مساوی ۳۴/۱ سانتی‌متر بودند که از این تعداد ۸۶/۶ ٪ متعلق به نژاد فارس بودند.

گره چهارم، زیر گروه دیگری از گره دوم با ۲۶۰ نفر که دارای پهنای شانه بین دو زائده آکرومیون بزرگتر از ۳۴/۳۱ سانتی‌متر بودند که از این تعداد ۴۶/۲٪ نژاد فارس و ۲۶/۹٪ ترک بودند.

جدول ۳- نتایج حاصل از ماتریس دوران یافته

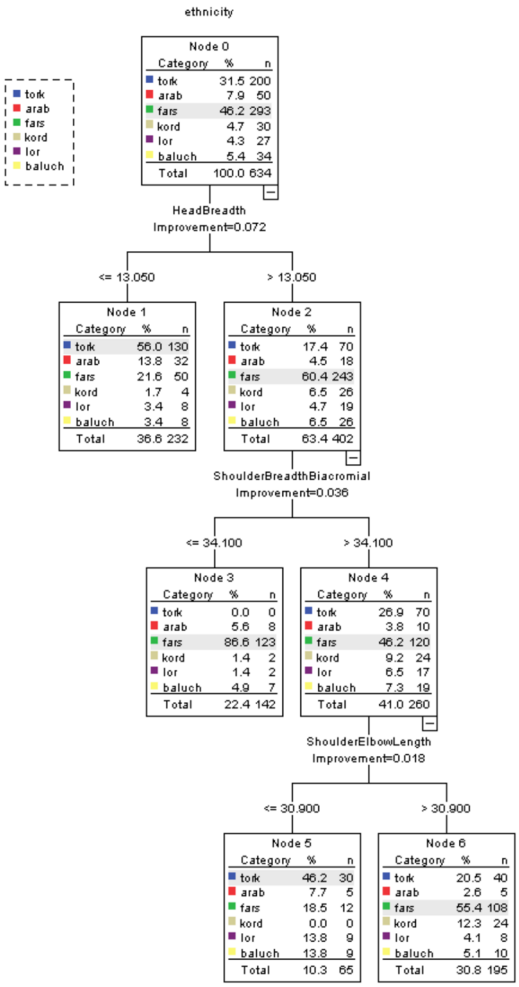
ابعاد آنتروپومتری	مولفه اول (PC_1)	مولفه دوم (PC_2)	مولفه سوم (PC_3)	مولفه چهارم (PC_4)	مولفه پنجم (PC_5)	مولفه ششم (PC_6)	مولفه هفتم (PC_7)
طول قد	۰/۴۷۳	۰/۴۸۱	۰/۶۱۷	۰/۰۱۰-	۰/۲۶۱	۰/۰۶۸	۰/۱۶۸
ارتفاع چشم	۰/۴۳۴	۰/۴۹۲	۰/۶۲۱	۰/۰۰۳	۰/۲۶۵	۰/۰۱۹	۰/۱۷۵
ارتفاع شانه	۰/۴۹۰	۰/۴۸۴	۰/۶۰۵	۰/۰۱۷	۰/۲۳۲	۰/۷۳۱	۰/۱۴۹
ارتفاع آرنج	۰/۳۸۱	۰/۵۷۷	۰/۵۸۹	۰/۰۱۰-	۰/۱۲۱	۰/۰۱۱	۰/۱۲۵
ارتفاع کفل	۰/۲۳۱	۰/۰۴۶	۰/۷۵۷	۰/۰۲۸	۰/۰۶۶	۰/۰۸۲	۰/۰۰۷-
ارتفاع برآمدگی انگشتان	۰/۱۷۱	۰/۴۶۹	۰/۷۰۰	۰/۰۷۰	۰/۱۴۷	۰/۰۱۳-	۰/۱۸۱
ارتفاع نوک انگشتان میانه	۰/۱۱۶	۰/۴۷۵	۰/۶۹۲	۰/۰۵۴	۰/۰۳۵	۰/۰۷۸-	۰/۱۴۳
ارتفاع نشسته	۰/۳۵۹	۰/۷۱۶	۰/۲۶۸	۰/۰۵۷	۰/۲۴۰	۰/۰۸۲-	۰/۱۲۲
ارتفاع چشم نشسته	۰/۲۵۷	۰/۷۵۸	۰/۲۷۸	۰/۰۶۵	۰/۲۳۸	۰/۰۸۳-	۰/۱۳۲
ارتفاع آرنج نشسته	۰/۳۳۷	۰/۷۴۹	۰/۲۱۶	۰/۱۴۶	۰/۱۳۲	۰/۰۹۴	۰/۰۶۰-
ارتفاع شانه نشسته	۰/۰۸۰-	۰/۸۵۴	۰/۰۱۰-	۰/۰۷۱	۰/۰۱۰-	۰/۱۹۴	۰/۰۲۳
ضخامت ران	۰/۰۷۵	۰/۰۸۲	۰/۱۰۴	۰/۷۳۶	۰/۱۰۷	۰/۱۱۰-	۰/۰۵۹-
طول کفل زانو	۰/۲۸۴	۰/۰۴۰-	۰/۶۴۹	۰/۳۰۱	۰/۲۱۷	۰/۰۵۶	۰/۰۱۶
طول کفل رکبی	۰/۱۷۲	۰/۰۸۸-	۰/۵۸۸	۰/۲۱۸	۰/۰۴۹	۰/۴۴۹	۰/۲۸۶-
ارتفاع زانو	۰/۳۸۵	۰/۳۷۶	۰/۴۳۵	۰/۰۷۹	۰/۴۰۳	۰/۲۰۲-	۰/۱۷۳
پهنای شانه(بین دو عضله دلتوئید)	۰/۰۶۵	۰/۲۲۹	۰/۰۹۲	۰/۵۶۷	۰/۳۰۱	۰/۲۳۶	۰/۴۲۴
پهنای شانه بین دو زائده آکرومیون	۰/۰۱۶	۰/۱۰۹	۰/۱۴۱	۰/۱۴۲	۰/۰۰۱-	۰/۰۲۵	۰/۸۷۱
پهنای باسن	۰/۰۲۱	۰/۰۲۳-	۰/۰۳۱	۰/۵۱۷	۰/۱۳۲-	۰/۴۵۷	۰/۳۸۳
عمق سینه	۰/۱۸۸-	۰/۱۰۱-	۰/۰۱۱-	۰/۷۷۰	۰/۱۰۰	۰/۰۹۸	۰/۰۳۳-
عمق شکم	۰/۰۴۸-	۰/۰۶۲	۰/۰۱۲-	۰/۷۸۷	۰/۰۵۴	۰/۲۵۱	۰/۰۴۶
طول شانه-آرنج	۰/۵۴۲	۰/۱۶۹	۰/۲۷۳	۰/۱۰۴-	۰/۳۰۸	۰/۰۵۹-	۰/۳۱۴
طول آرنج-نوک انگشتان	۰/۷۴۱	۰/۱۱۱	۰/۲۲۵	۰/۱۰۹	۰/۰۵۶-	۰/۲۶۳	۰/۱۷۹-
طول اندام فوقانی	۰/۸۰۷	۰/۲۰۵	۰/۲۴۵	۰/۰۵۷-	۰/۱۵۱	۰/۰۵۱	۰/۰۱۰
طول شانه-چنگش	۰/۸۱۸	۰/۲۰۸	۰/۱۵۹	۰/۰۱۳-	۰/۱۱۸	۰/۰۷۴-	۰/۱۲۰-
طول سر	۰/۱۱۴-	۰/۰۹۸	۰/۰۸۱	۰/۱۳۹	۰/۲۷۶	۰/۷۰۱	۰/۰۲۴
پهنای سر	۰/۲۰۸	۰/۳۹۴	۰/۱۱۴-	۰/۲۰۳	۰/۳۱۱	۰/۴۰۶	۰/۰۲۰-
طول دست	۰/۳۴۷	۰/۱۶۷	۰/۲۸۸	۰/۰۰۸	۰/۶۰۷	۰/۲۶۷	۰/۰۵۳
پهنای دست	۰/۰۷۷	۰/۲۲۰	۰/۱۱۴	۰/۲۲۲	۰/۷۲۱	۰/۱۷۵	۰/۱۷۲
طول کف پا	۰/۴۵۲	۰/۲۰۹	۰/۲۸۰	۰/۰۸۸	۰/۵۹۸	۰/۰۳۸-	۰/۰۴۴-
پهنای کف پا	۰/۱۳۴	۰/۱۱۱	۰/۱۰۸	۰/۳۷۷	۰/۶۵۷	۰/۱۰۸	۰/۲۹۲-
فاصله بین نوک انگشتان میانی دست راست و چپ زمانی که بازوها کاملا باز باشند	۰/۶۳۹	۰/۲۹۴	۰/۳۹۹	۰/۰۲۴	۰/۳۵۰	۰/۰۳۸-	۰/۱۱۹
فاصله بین آرنج ست راست و چپ هنگامی که بازوها به طرفین باز شده و ساعدها خم شده	۰/۶۹۳	۰/۰۴۹	۰/۲۵۶	۰/۰۳۳-	۰/۳۰۴	۰/۲۴۶-	۰/۱۵۳
حد دسترسی چنگش ایستاده	۰/۵۶۹	۰/۴۸۱	۰/۴۱۴	۰/۰۱۰	۰/۲۶۵	۰/۲۰۹	۰/۰۸۶
حد دسترسی چنگش نشسته	۰/۳۹۲	۰/۵۸۶	۰/۱۸۷	۰/۱۰۶	۰/۲۴۲	۰/۴۰۲	۰/۰۱۹-
حد دسترسی چنگش به جلو	۰/۶۳۷	۰/۱۷۱	۰/۰۶۷	۰/۰۷۷	۰/۰۷۸-	۰/۵۴۶	۰/۱۳۷
وزن	۰/۱۶۳	۰/۲۹۳	۰/۲۲۹	۰/۶۸۸	۰/۱۱۲	۰/۰۲۹-	۰/۱۹۴
نژاد	۰/۱۱۱-	۰/۰۲۰-	۰/۰۵۱.	۰/۰۳۸	۰/۰۸۳-	۰/۰۶۵.	۰/۸۵۴

گره پنجم و ششم زیر گروهی از گره چهارم بر اساس طول آرنج-نوک انگشتان هستند. گره پنجم مربوط به طول‌آرنج-نوک انگشتان کوچکتر یا مساوی ۳۰/۹ سانتی‌متر و گره ششم مربوط به طول‌آرنج-نوک - انگشتان بزرگتر از ۳۰/۹ سانتی‌متر می‌باشد. گره پنجم شامل ۶۵ نفر بود که از این تعداد ۴۶/۲٪ به نژاد ترک و ۱۸٪/۵ به نژاد فارس تعلق داشت. برای گره ششم ۵۵٪/۴ از ۱۹۵ نفر از افراد را نژاد فارس و ۲۰/۴٪ را ترک تشکیل دادند.

براساس نتایج بدست آمده از الگوریتم درخت تصمیم

جدول ۴ – قوانین استخراجی از درخت تصمیم برای زنان

قانون	تعداد افراد	درصد
پهنای صورت ≥ ۱۳,۰۵	۲۳۲	۳۶,۶
پهنای صورت ≤ ۱۳,۰۵	۱۴۲	۲۲,۴
پهنای شانه بین دو زائده آکرومیون ≥ ۳۴,۱		
پهنای صورت ≤ ۱۳,۰۵	۱۹۵	۳۰,۸
پهنای شانه بین دو زائده آکرومیون < ۳۴,۱		
طول آرنج–نوک انگشتان < ۳۰,۹		
پهنای صورت ≤ ۱۳,۰۵	۶۵	۱۰,۳
پهنای شانه بین دو زائده آکرومیون < ۳۴,۱		
طول آرنج–نوک انگشتان ≥ ۳۰,۹		

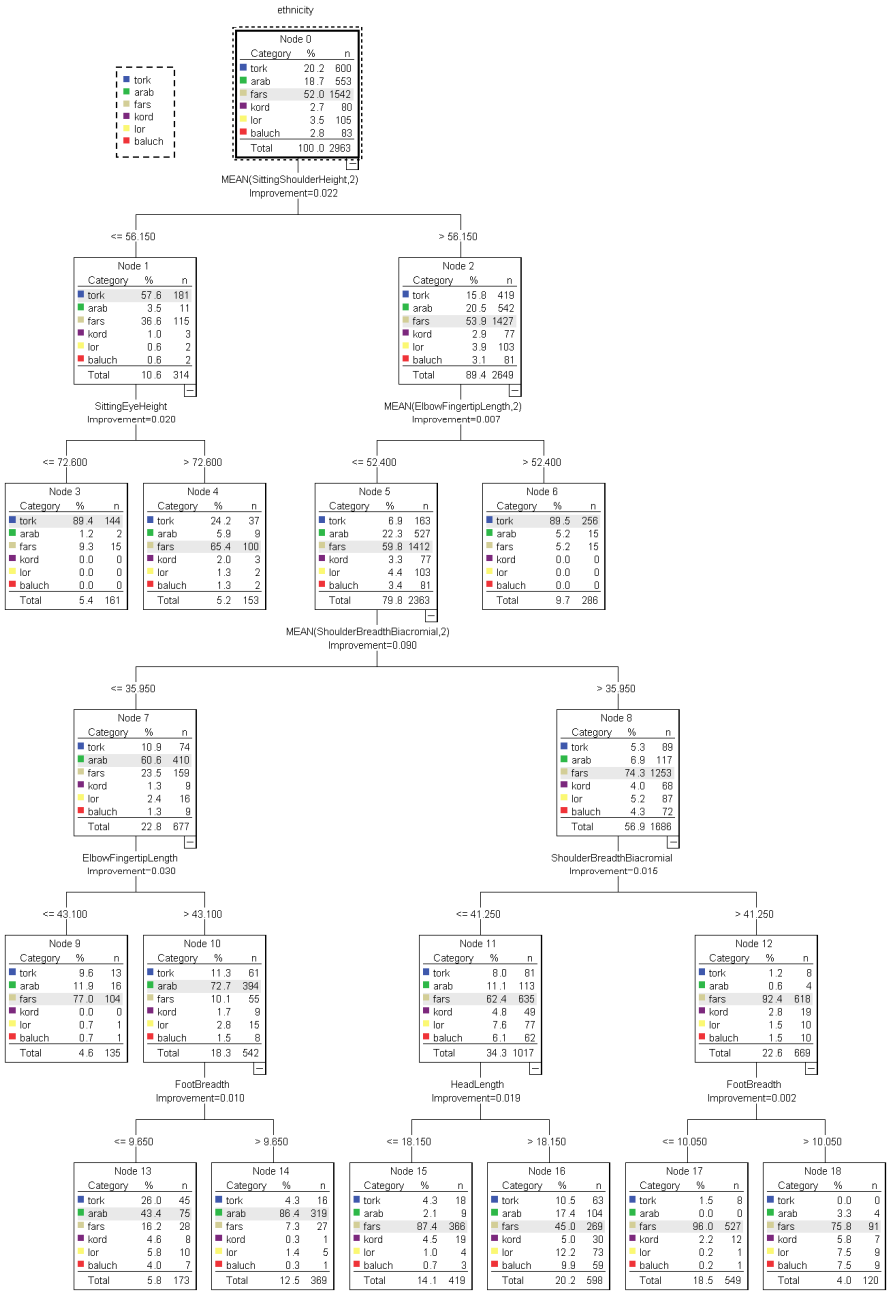


شکل ۲– نتایج درخت تصمیم گیری برای زنان

در مردان شش نژاد مورد مطالعه (شکل ۳)، در کل بین ۲۹۶۳ نفر روابط برقرار شد، که از این تعداد ۵۲٪ نژاد فارس، ۲۰/۲٪ نژاد ترک، ۱۸/۷٪ نژاد عرب، ۳/۵٪ نژاد لر، ۲/۸٪ نژاد بلوچ و ۲/۷٪ را نژاد کرد به خودشان اختصاص داده‌اند.

براساس این نمودار، گروه مردان با استفاده از ۱۸ گره به هم مرتبط شده‌اند.

گره اول براساس ارتفاع‌شانه‌نشسته با افرادی با اندازه کوچکتر یا مساوی ۵۶/۱۵ سانتی‌متر بوده که ۳۱۴ نفر را شامل شده است. از این تعداد ۵۷/۶٪ ترک و ۳۶/۶٪



شکل ۳- نتایج درخت تصمیم گیری برای مردان

فارس را شامل می‌شوند.

گره سوم و چهارم با متغیر تن-سنجی ارتفاع-نشسته-چشم زیر گروهی از گره اول می‌باشند. بطوری‌که مقادیر کمتر یا مساوی ۷۲/۶ سانتی‌متر برای ارتفاع نشسته-چشم گره سوم را تشکیل داده که نژاد ترک ۸۹/۴٪ از ۱۶۱ نفر از کل افراد گروه سوم را به خود اختصاص داده‌اند. درگره چهارم هم با ۱۵۳ نفر، ۶۵/۴٪ از آن مربوط به نژاد فارس و ۲۴/۲٪ از افراد گره چهارم مربوط به نژاد ترک می‌باشند.

زنجیره سوم از این نمودار که از سه گره فراتر نرفته است مربوط به گره صفر، گره دوم و گره شش می‌باشد. گره دوم با عامل تن‌سنجی ارتفاع‌شانه نشسته بیشتر از ۵۶،۱۵ سانتی‌متر و گره ششم با طول‌آرنج-تا نوک-انگشتان بیشتر از ۵۲،۴ سانتی‌متر از زیر گره‌های زنجیره سوم می‌باشند.

زنجیره دیگر با ۲۳۶۳ نفر از ۵ تا ۶ گره تشکیل شده

جدول ۵- قوانین استخراجی درخت تصمیم گیری برای مردان

قانون	تعداد افراد کلاس	درصد افراد کلاس
ارتفاع شانه نشسته ≥ ۵۶,۱۵ ارتفاع دید نشسته < ۷۲,۶	۱۵۳	۵,۲
ارتفاع شانه نشسته ≥ ۵۶,۱۵ ارتفاع دید نشسته ≥ ۷۲,۶	۱۶۱	۵,۴
ارتفاع شانه نشسته < ۵۶,۱۵ طول آرنج-تا نوک انگشتان < ۵۲,۴	۲۸۸	۹,۷
ارتفاع شانه نشسته < ۵۶,۱۵ طول آرنج-تا نوک انگشتان ≥ ۵۲,۴ پهنای شانه بین دو زائده آکرومیون < ۴۱,۲۵ پهنای کف پا < ۱۰,۰۵	۱۲۰	۴
ارتفاع شانه نشسته < ۵۶,۱۵ طول آرنج-تا نوک انگشتان ≥ ۵۲,۴ پهنای شانه بین دو زائده آکرومیون < ۴۱,۲۵ پهنای کف پا ≥ ۱۰,۰۵	۵۴۹	۱۸,۵
ارتفاع شانه نشسته < ۵۶,۱۵ طول آرنج-تا نوک انگشتان ≥ ۵۲,۴ پهنای شانه بین دو زائده آکرومیون ≥ ۴۱,۲۵ طول سر < ۱۸,۱۵	۵۹۸	۲۰,۲
ارتفاع شانه نشسته < ۵۶,۱۵ طول آرنج-تا نوک انگشتان ≥ ۵۲,۴ پهنای شانه بین دو زائده آکرومیون ≥ ۴۱,۲۵ طول سر ≥ ۱۸,۱۵	۴۱۹	۱۴,۱
ارتفاع شانه نشسته < ۵۶,۱۵ طول آرنج-تا نوک انگشتان < ۴۳,۱ پهنای شانه بین دو زائده آکرومیون ≥ ۳۵,۹۵ عرض پا < ۹,۶۵	۳۶۹	۱۲,۵
ارتفاع شانه نشسته < ۵۶,۱۵ طول آرنج-تا نوک انگشتان < ۴۳,۱ پهنای شانه بین دو زائده آکرومیون ≥ ۳۵,۹۵ عرض پا ≥ ۹,۶۵	۱۷۳	۵,۸

است. که خود این زنجیره از هفت زنجیره مجزا تشکیل شده است. که بطور خلاصه به قرار زیر می‌باشند:

- گره صفر، گره دوم، گره پنجم، گره هفتم و گره نهم
- گره صفر، گره دوم، گره پنجم، گره هفتم، گره دهم و گره سیزدهم
- گره صفر، گره دوم، گره پنجم، گره هفتم، گره دهم و گره چهاردهم
- گره صفر، گره دوم، گره پنجم، گره هشتم، گره یازدهم و گره پانزدهم
- گره صفر، گره دوم، گره پنجم، گره هشتم، گره یازدهم و گره شانزدهم
- گره صفر، گره دوم، گره پنجم، گره هشتم، گره دوازدهم و گره هفدهم
- گره صفر، گره دوم، گره پنجم، گره هشتم، گره دوازدهم و گره هجدهم

بحث و نتیجه‌گیری

امروزه جمع‌آوری داده‌های مربوط به ابعاد بدن از گروه‌ها، جمعیت‌ها و نژادهای مختلف در علوم مختلف بویژه ارگونومی اهمیت بسیاری داشته و در مراکز تحقیقاتی مختلف نیز با مقاصد متفاوت داده‌ها گردآوری می‌شوند (۳۱-۳۳). تغییر شکل داده‌های خام به

اطلاعات معنی‌دار یکی از مهمترین موضوعات چالش برانگیز در حوزه های مختلف از جمله حوزه سلامت می‌باشد (۳۱). تولید مستمر انبوهی از داده‌ها (۳۴) از طریق مطالعه بر روی این داده‌ها و کشف روابط بین آنها می‌تواند یکی از اهداف جمع‌آوری این داده‌ها باشد.

بانک‌های اطلاعاتی پایه و اساس طراحی ارگونومیکی محیط‌های‌کاری، ایستگاه‌های‌کاری و ابزار و وسایل تولیدی می‌باشند. توسعه این بانک‌ها نسبت به دهه‌های گذشته (۳۵)، نیازهای جدیدی از قبیل اندازه-گیری دقیق ابعاد بدن، بروزکردن داده‌ها، خلاصه‌سازی خودکار داده‌ها و استخراج اطلاعات ذخیره شده و کشف الگوها از داده های خام حائز اهمیت است.

با پیشرفت تکنولوژی و به کارگیری علوم نوین، محققان سعی در پیدا کردن جایگزین‌هایی برای روش‌های آماری خطی با مدل‌ها با روابط کارکردی متفاوت هستند. در این رابطه داده‌کاوی بعنوان، یک روش جایگزین برای برآورد اندازه‌گیری‌های تن‌سنجی مطرح گردیده است (۳۶). از دلایل اصلی استفاده از داده‌کاوی، میتوان به رویارویی با مجموع عظیم داده‌ها اشاره نمود (۳۵)، که کنارآمدن محققان با حجم انبوهی از داده‌ها با اتکا به شیوه‌های سنتی غیرممکن می‌باشد (۳۹-۳۸).

استخراج اطلاعات و دانش از داده‌ها مفهوم دیرینه‌ای در مطالعات علمی داشته و آنچه که جدید می‌باشد همگرایی و اشتراک چندین رشته و فن‌آوری‌های متناظر آنها است که فرصت منحصر به فردی را برای داده کاوی ایجاد کرده است (۴۰). داده‌کاوی، استخراج دانش از اطلاعات ذخیره شده در پایگاه داده‌ها و ایجاد شرح روشن و قابل فهم از الگوها است که بعنوان ابزاری برای تحلیل داده‌های وسیع و مدل‌سازی پیشگویانه با روش‌های محاسباتی جدید استفاده می‌شود. داده‌کاوی در طول چند سال گذشته، با موفقیت در طیف وسیعی از حوزه‌های مختلف بکار گرفته شده (۳۲) و نتیجه این کاربرد منجر به حل، پیش‌بینی و طبقه‌بندی داده‌ها شده است (۴۱-۳۴). علاوه بر کاربرد آن در آنتروپومتری بعنوان مثال در صنعت پوشاک (۴۴) دسته‌بندی سربازان (۴۵)، می‌توان به کاربردهای آن در حیطه پزشکی (۳۹) از جمله استخراج عوامل موثر در بیماریهای قلبی-عروقی، پیش‌بینی عوارض بیماری عروق‌کرونری، پیش‌بینی عوارض بیماران دیابتی، کنترل عفونت‌های بیمارستانی، و کاربرد آن در بیماران مبتلا به آسم (۴۶) اشاره نمود. همچنین به استخراج روابطی برای تشخیص جنسیت از روی داده‌های آنتروپومتریک حاصل از فیلمبرداری اشاره کرد (۴۷).

تشخیص تاثیر تفاوت‌های نژادی در آنترپومتری، از موضوعات مورد توجه پژوهشگران و طراحان و صنایع تولیدی مختلف می‌باشد. لذا در مطالعه حاضر برای کشف روابط بین متغیرها با نژاد از دسته‌بندی داده‌ها با استفاده از الگوریتم‌های درخت تصمیم استفاده شد. از آنجائیکه دقت عملیات داده‌کاوی برای تجزیه و تحلیل، بستگی به انتخاب مناسب متغیر ورودی (داده‌ها و ابعاد) دارد و علاوه بر آن، بعلت زیاد بودن داده‌های تن سنجی (۱۳۷۶۴۰=۳۷*۳۷۲۰)، زمان و فضای بیشتری برای پردازش داده‌ها لازم می‌باشد (۴۸). در مطالعه حاضر ۳۷ متغیر ورودی برگرفته از ۳۷۲۰ نفر از نژادهای مختلف مورد بررسی قرار گرفت.

قبل از ورود به استخراج قوانین با استفاده از روش تجزیه و تحلیل عامل‌های اصلی (PCA)، داده‌های مورد بررسی به داده‌های مستقل کاهش داده شده (۴۸) و سپس مراحل داده‌کاوی و درخت تصمیم پیاده گردید. معیار انتخابی در روش تجزیه و تحلیل عامل‌های اصلی، براساس نتایج مطالعات پیشین ۰/۶۵ انتخاب گردید (۴۹). با به کارگیری معیار انتخابی فوق، از ۳۷ متغیر ورودی، تنها ۲۱ متغیر ضریب بالای ۰/۶۵را بدست آورده و در هفت خوشه قرار گرفتند.

عامل‌های مورد بررسی شامل ارتفاع شانه، ارتفاع کفل، ارتفاع برآمدگی انگشتان، ارتفاع نوک از انگشت میانه، ارتفاع نشسته، ارتفاع چشم نشسته، ارتفاع آرنج نشسته، ارتفاع شانه نشسته، ضخامت ران، پهنای شانه بین دو زائده آکرومیون، عمق سینه، عمق شکم، طول آرنج-نوک انگشتان، طول اندام فوقانی، طول شانه-چنگش، طول سر، پهنای دست، پهنای کف پا، فاصله بین آرنج دست راست و چپ هنگامی که بازوها به طرفین باز شده و ساعدها خم شده و نژاد بودند.

یافته‌های بدست آمده از مطالعه حاضر نشان داد که در هر دو گروه جنسیتی (مردان و زنان) مورد مطالعه اکثریت جامعه را نژاد فارس بخود اختصاص داده است. در گروه مردان، نژاد فارس در یازده گره، و نژادهای ترک و عرب نیز هر کدام در سه گره عامل اکثریت (تعیین‌کننده) بودند. در حالیکه در گروه زنان، نژاد

فارس در چهار گره و در نژاد ترک در دو گره عامل اکثریت (تعیین‌کننده) بودند. تحلیل این یافته‌ها نشان داد که در میان مردان، ارتفاع شانه نشسته، ارتفاع دید نشسته، طول‌آرنج-نوک انگشتان، پهنای شانه بین دو زائده آکرومیون، عرض پا و طول سر به ترتیب از عامل‌های متمایزکننده در دسته‌بندی و ایجاد تفاوت‌های نژادی در گروه‌های مورد مطالعه بودند. درحالیکه در میان زنان عامل‌های پهنای‌صورت، پهنای شانه بین دو زائده آکرومیون و طول آرنج-نوک انگشتان از عامل‌های متمایزکننده در دسته‌بندی و باعث تفاوت‌های نژادی می‌باشند.

تشخیص تاثیر تفاوت‌های نژادی در آنتروپومتری، از موضوعات مورد توجه پژوهشگران و طراحان و صنایع تولیدی مختلف بوده و یکی از یافته‌های مطالعه حاضر نیز مشخص شدن عوامل متمایزکننده در دسته‌بندی و ایجاد تفاوت‌های نژادی در نژادهای مختلف مورد مطالعه است. از طرف دیگر طراحان در هنگام طراحی ایستگاه‌های کاری، ابزارهای دستی و ماشین‌الات علاوه بر الزامات کار، بایستی ویژگی‌های آناتومیکی، فیزیولوژیکی و آنتروپومتریکی کاربران را نیز مد نظر قرار دهند و این ابعاد (ارتفاع دید نشسته، طول‌آرنج-نوک انگشتان، پهنای شانه بین دو زائده آکرومیون، و عرض پا) نیز از مهمترین پارمترهای تاثیرگذار در طراحی بویژه ایستگاه‌های کاری می‌باشد. لذا می‌توان از این یافته در طراحی بهینه ایستگاه‌های‌کاری ارگونومیک برای مردان کارگر ایرانی با نژادهای مختلف استفاده نمود.

همچنین با در نظر گرفتن هدف مطالعه و نتایج حاصل از درخت تصمیم، می‌توان نتیجه گرفت که فارغ از نژاد، عامل‌های تمایزکننده در مردان مربوط به عامل‌های تن‌سنجی (هیکل) و عامل‌های تمایز کننده در زنان مربوط به عامل‌های تن‌سنجی صورت (زیبایی) می‌باشند. نتایج مطالعه حاضر همچنین می‌تواند تاکید مجددی بر سودمندی بکارگیری روش درخت تصمیم برای بررسی تداخلات بین متغیرهای پیشگویی‌کننده باشد که در آن با ترکیب تنوع گره‌ها در ساختار کلی، مهمترین عوامل تعیین‌کننده نژادی را می‌توان شناسایی نمود. همچنین نتایج مطالعه حاضر تائید نمود که درخت تصمیم می‌تواند بعنوان یک روش موثر در شناسایی متغیرهای مهم برای دسته‌بندی و شناخت تفاوتهای نژادی در حیطه آنتروپومتری بکار گرفته شود.

در نهایت براساس نتایج مطالعه حاضر می‌توان نتیجه گرفت که عوامل متمایز کننده در مردان و زنان جامعه ایرانی متفاوت از هم بوده و این یافته می‌تواند علاوه برکاربرد در صنایع تولیدی و طراحی، در موارد دیگری همچون تشخیص هویت در پزشکی قانونی، ساخت محصولات ارتوپدی نیز کاربرد داشته باشد.

تقدیر و تشکر

مولفین صمیمانه از معاونت محترم دانشگاه علوم پزشکی تبریز به دلیل تامین مالی این مطالعه از قبل طرح تحقیقاتی ۲۹۰/ب تشکر و قدردانی می‌کنند و همچنین از همکاری آقای دکتر فردوسی بایرامی کمال تشکر و قدردانی را داریم.

References

1. Phisant E. Human, anthropometry, ergonomics and design. Trans Choobineh AR, Mououdi MA 3rd ed Tehran: Maad Press. 2012:11.

2. Heydari P, Jafarvand M, Alizadeh SS, Varmazyar S, Hakimi C. Predicting maximum oxygen consumption based on anthropometric dimensions in medical emergency students. Iran Occup Health. 2018;14(6):126-134. [In Persian]

3. Ministry of health and medical education, Environmental and Occupational health center, control of occupational factors affecting health department, Occupational Health Statistics Unit, 2010.

4. Moudy AM. Static anthropometric characteristics determine the design and evaluation of seat comfort saddle promised the faithful. Iran Occup Health. 2013;4:9-24.

5. Jiao JR, Zhang Y, Helander M. A Kansei mining system for affective design. Expert Systems Applicat. 2006;30(4):658-73.

6.Mououdi MA. The determination of static anthropometry characteristics for designing and evaluating the comfort of saddle chair. Iran Occup Health. 2013;9(4):24-29. [In Persian]

7.Habibi E, Sadeghi N, Mansouri F, Sadeghi MRM, Ranjbar M. Comparison of Iranian student's anthropometric information and American and English standards. J Jahrom Uni Med Sci. 2012;10(2):22-30.

8. Kazem Haghighi M, Saremi M. The situation of anthropometric databank in Iran: A review study. Iran Occup Health. 2017;14(4):106-117. [In Persian]

9.Lewin T. Anthropometric studies on Swedish industrial workers when standing and sitting. Ergonomics. 1464;12(6):4.25-884

10.Krishan K, Kanchan T, Passi N. Estimation of stature from the foot and its segments in a sub-adult female population of North India. J Foot Ankle Res. 2011;4(1):24.

11.Chung MJ, Lin HF, Wang MJJ. The development of sizing systems for Taiwanese elementary-and high-school students. Int J Indust Ergonom. 2007;37(8):707-16.

12. Abledu J, Abledu G, Offei E, Antwi E. Estimation of stature and body weight from footprint dimensions among a female population in Ghana. Australian J Forensic Sci. 2015:1-8.

13. Agha SR, Alnahhal MJ. Neural network and multiple linear regression to predict school children dimensions for ergonomic school furniture design. Appl Ergonom. 2012;43(6):979-84.

14. Hsu CH, Chen SC, Liu BS, editors. Employing data mining to identify the significant rules for classifying body types. Proceedings of the 2007 annual Conference on International Conference on Computer Engineering and Applications; 2007: World Scientific and Engineering Academy and Society (WSEAS).

15. Pheasant S, Haslegrave CM. Bodyspace: Anthropometry, ergonomics and the design of work: CRC Press; 2016.

16. Woodson WE, Tillman B, Tillman P. Human factors design handbook: information and guidelines for the design of systems, facilities, equipment, and products for human use.1992.

17. Sadeghi F, Mazloumi A, Kazemi Z. Study of anthropometric measurements of Iranian workers in factories Persian ethnicity Tehran, Isfahan and Fars. Occup Med Quart J. 2013; 5(1):34-45.

18. M AE. Body mechanics and workstation design principles (ergonomics). Tehran: Omide Majd Publ. 2000.

19. Ashoori M, Naji Moghaddam V, Alizadeh S, Safi M. Classification and Clustering Algorithm Application for Prediction of Tablet Numbers: Case Study Diabetes Disease. Health Information Management. 1392;10(5). [Persian]

20. Rashidmehrabadi A, Pedram M. Classification and identification of blood donors tend to donate blood in the future. The Fourth Iran Data Mining Conference, Sharif University of Technology. 2010, Tehran.

21. Tan J. Medical Informatics: Concepts, Methodologies, Tools, and Applications. Hershey: IGI Global snippet; 2008.

22. Koh HC, Tan G. Data mining applications in healthcare. J Healthc Inf Manag 2005; 19(2): 64-72.

23. Phillips KT. Design and testing of a knowledge-based system for treatment of heart failure patients in a hospital system: University of Iowa; 2005.

24. Pyle D. Data preparation for data mining: Morgan Kaufmann; 1999.

25. Lin HF, Hsu CH, Wang MJJ, Lin YC. An application of data mining technique in developing sizing system for army soldiers in Taiwan. WSEAS Transactions on Computers. 2008;7(4):245-52.

26. Makooie FM, Hadizadeh M, Payvandy P. Applying Self-organizing Neural Network Based on Data Mining to Establish a Clothing Size Chart. Textile Sci Technol. 2012;2(2):97-103.

27. Park JK. Classification of foot types for shoes size system of elderly women. J Korean Soc Costume. 2005;55(2):33-44.

28. Sadeghi F, Mazloumi A, Kazemi Z. An anthropometric data bank for the Iranian working population with ethnic diversity. Appl Ergonom. 2015; 48:95-103

29. Chen H, Fuller SS, Friedman C, Hersh W. Medical Informatics: Knowledge Management And Data Mining in Biomedicine. New York: Springer; 2005.

30. Gupta G.K,Introduction to Data Miningwith Case Studies, PrenticeHall of India, second edition, 2011.

31. Balib RK. Clinical Knowledge Management: Opportunities and Challenges. Hershey: Idea Group Inc (IGI); 2005

32. Canlas RD. Data Mining in Healthcare: Current Applications and Issues. Carnegie Mellon University, Australia. 2009.

33. Berka P, Rauch J, Zighed DA. Data Mining and Medical Knowledge Management: Cases and Applications. Hershey: Idea Group Inc (IGI); 2009.

34. Zhu L, Wu B, Cao C. Introduction to medical data mining. Sheng Wu Yi Xue Gong Cheng Xue Za Zhi. 2003;20(3):559-562.

35. Kaur H, Wasan SK. Empirical Study on Applications of Data Mining Techniques in Healthcare. J Comput Sci. 2006;2(2): 194-200

36.Dursun Kaya M, Samet Hasiloglu A, Bayramoglu M, Yesilyurt H, Fahri Ozok A. A new approach to estimate anthropometric measurements by adaptive neuro-fuzzy inference system. Int J Indust Ergonom. 2..4;42(2):135-1.7

37. Wright A, Chen ES, Maloney FL. An automated technique for identifying associations between medications, laboratory results and problems. J Biomed Inform. 2010;43(6):891-901.

38. Moghaddassi H, Hoseini A, Asadi F, Jahanbakhsh M. Application of Data Mining. Health Inform Manag. 2012;9(2):304. [In Persian]

39. Gholamhosseini L, Damroodi M. Evaluation of Data Mining Applications in the Health System. Paramed Sci Military Health. 2015;10(1):39-48 [In Persian]

40. Moghaddasi H. Information Quality in Health care. Tehran: Word Processing Publishing; 2005:41. [In Persian]

41. Han J, Kamber M, Pei J. Data mining, southeast asia edition: Concepts and techniques: Morgan kaufmann; 2006.
42. Hsu CH, Liu BS, Chen SC, editors. Using data mining to extract sizing knowledge for promoting manufacture. Proceedings of the 6th WSEAS international conference on Applied computer science; 2006: World Scientific and Engineering Academy and Society (WSEAS)
43. Jiao JR, Zhang Y, Helander M. A Kansei mining system for affective design. Expert Systems Applicat. 2006;30: 658-673.
44. Elfaki Elnour EF, Ali AHM, Elhafiz M, Mohammed SA. Application of data mining methods in establishing sizing system for clothing industry. Int J Engineer Sci Res Technol. 2015;4(8):32-39.
45. Lin HF, Hsu CH, Wang MJ, Lin YC. 'An application of data mining technique in developing sizing system for army soldiers in Taiwan' World Scientific and Engineering and Society (WSEAS). Transact Comput. 2008;7:245-252.
46 Samad Soltani T, Langarizadeh M, Zolnoori M. Data Mining and Analysis: Reporting Results for Patients With Asthma. J Payavard Salamat. 2015;9(3):224-234 [In Persian]
47. Camalan S, Sengul G, Misra S, Maskeli R, Damaevi R. "Gender Detection using 3D anthropometric measurements by Kinect," Metrol Measur Systems. 2018;25(2): 253-267.
48. Usman Ali M, Ahmed Sh, Ferzund J, Mehmood A, Rehman A. Using PCA and Factor Analysis for Dimensionality Reduction of Bio-informatics Data. Int J Adv Comput Sci Applicat. 2017;8(5):415-426
49. Palaniandy T, Yahya MS, Zainun NY. Development of Anthropometric Model Using Artificial Neural Networks (ANN) Approach. Int J Construct Technol Manag. 2013;1(1):1-5.