Scientific Journal of
Rehabilitation Medicine

May and June 2022. Volume 11. Number 2

Research Paper

Comparing the Effectiveness of Two Simple Decision Tree and Random Forest Algorithms in Predicting Osteoporosis in Active Middle-Aged Men

*Leila Fasihi[1], Bakhtyar Tartibian[1], Rasoul Eslami[1]

1. Department of Exercise Physiology, Faculty of Physical Education and Sport Sciences, Allameh Tabataba'i University, Tehran, Iran.

Use your device to scan and read the article online

Citation Fasihi L, Tartibian B, Eslami R. [Comparing the Effectiveness of Two Simple Decision Tree and Random Forest Algorithms in Predicting Osteoporosis in Active Middle-Aged Men (Persian)]. Scientific Journal of Rehabilitation Medicine. 2022; 11(2):264-275. https://dx.doi.org/10.32598/SJRM.11.2.10

https://dx.doi.org/10.32598/SJRM.11.2.10

This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0)

Received: 15 Mar 2021
Accepted: 19 Apr 2021
Available Online: 22 May 2022

ABSTRACT

Background and Aims Osteoporosis is called a "silent disease" because it has no symptoms until a bone is fractured. Therefore, its early detection before occurrence of fracture is important. Using data mining algorithms can help access the information hidden in the data. This study aims to compare two simple decision tree and random forest algorithms to predict osteoporosis in active middle-aged men.

Methods A total of 256 middle-aged men referred to Ayatollah Kashani Hospital in Tehran, Iran during 2017-2020 participated in this study. Data analysis was carried out in MATLAB software version 2020. Evaluation was performed using the confusion matrix and based on accuracy and precision criteria.

Results Out of 103 factors related to personal information, lifestyle, and disease, 11 were selected as inputs to the algorithms. The results showed that the random forest algorithm had a better performance (73.4% accuracy and 68.07% precision) compared to simple decision tree.

Conclusion The data mining algorithms can be effective in predicting osteoporosis in active middle-aged men. These algorithms can be used for early treatment and rapid diagnosis of osteoporosis and prevent the occurrence of bone fractures and their irreparable complications.

Keywords Osteoporosis, Prediction, Data mining

* *Corresponding Author:*
Leila Fasihi
Address: Department of Exercise Physiology, Faculty of Physical Education and Sport Sciences, Allameh Tabataba'i University, Tehran, Iran.
Tel: +98 (918) 2183604
E-Mail: l_fasihi@atu.ac.ir

Extended Abstract

Introduction

Osteoporosis is a disease that weakens the bones and cause bone atrophy, and often affects people in their middle and later years. The disease begins at young age and progresses slowly over time. Today, osteoporosis is considered as a major threat in the world; its annual mortality rate is higher compared to all types of cancers. A study conducted by the Iranian Ministry of Health showed that in Iran, 44% of men and 47% of women aged >50 years have a low bone density, and 4.6% of people aged 20-70 years have osteoporosis in the spine. Globally, 1 in 3 women and 1 in 5 men over the age of 50 have osteoporosis.

The bone density is measured by a dual-energy x-ray absorptiometry (DEXA) method. In this method, the mean bone mineral density of patients is compared with the mean bone mineral density of young healthy people, which is called "T-score". There are several factors involved in causing osteoporosis, including uncontrollable factors such as age, gender, race, family history, menopause, and controlled factors such as lack of exercise, calcium and vitamin D deficiency in the diet, and high coffee, salt, cigarettes and alcohol consumption. Exercise is a factor that maintains and stimulates the bone formation, which reduces the risk of bone fractures through the accumulation of substances and minerals, strengthening muscles, and improving balance. The effect of physical activity and the resulting mechanical pressure on increasing bone density has been proven; the mechanical pressure on the bones through tendons and muscles has a direct effect on bone formation and deformation.

The science that helps us gain knowledge about data is called data mining. The power of data mining for the satisfactory diagnosis of various diseases has been proven. Data mining algorithms such as decision tree and random forest have been used successfully in various fields including medicine and sports. The use of data mining algorithms cause new information and relationships embedded in large and complex data become evident through inferring new relationships and learning patterns. In medicine, different tools are needed to predict the prevalence of osteoporosis in men and women. Data mining algorithms are a simple and efficient tool for predicting the risk of osteoporosis. With the help of these algorithms, it is possible to detect the risk of this disease without the need for diagnostic methods, and to identify the factors affecting this disease. This study aims to compare two simple decision tree and random forest algorithms to predict osteoporosis in active middle-aged men.

Materials and Methods

The study population consisted of 256 active men referred to Ayatollah Kashani Hospital in Tehran and had medical records. Of these, 89 were healthy, 126 had osteopenia and 41 were with osteoporosis. In this study, active men were those who had regular physical activity (three times a week) for at least one year. After signing a consent form, their DEXA information was used.

Two algorithms of simple decision tree and random forest were used to predict and classify healthy individuals, and patients with osteopenia and osteoporosis. Selected characteristics for algorithm training included: Age, height, weight, body mass index (BMI), physical activity, smoking, vitamin D and calcium intake, history of bone fracture, history of osteoporosis in the first-degree relatives, spinal curvature, and low back pain. For optimal use of data, it should be changed to be suitable for data mining algorithms. For the questions with "yes" and "no" answers, 1 and 0 were used, respectively. Then, the data was imported to MATLAB software version 2020 in an EXE file and analyzed. The algorithms were compared based on accuracy and precision criteria.

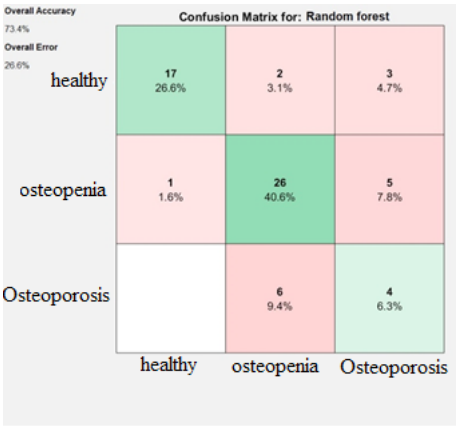
Results

Table 1 presents the descriptive information of participants.

Of collected data, 25% were used for testing and 75% for algorithm training. Figures 1 and 2 show the confusion ma-

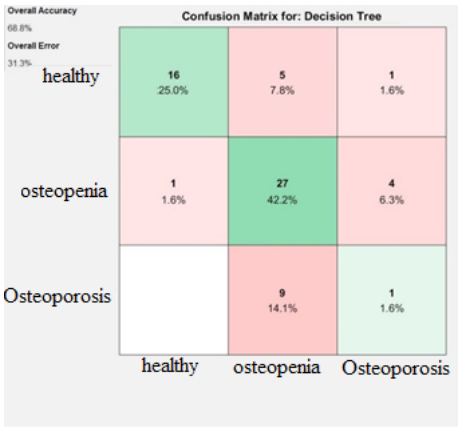
Table 1. Anthropometric characteristicsof the participants (n =256)

Characteristics	Mean± SD
Age (y)	57.30±7.67
Height (cm)	171.87±8.32
Weight (kg)	77.79±12.62
BMI (kg/m^2)	26.37±4.17



Scientific Journal of
Rehabilitation Medicine

Figure 1. Confusion matrix of random forest algorithm



Scientific Journal of
Rehabilitation Medicine

Figure 2. Confusion matrix of simple decision tree algorithm

trix of the two algorithms. It can be seen that the random forest algorithm had a better performance with 73.4% accuracy and 68.07% precision.

Discussion

Bone mineral density increases in childhood and reaches its maximum during puberty. After the third decade of life, general bone loss begins with aging, and bone strength decreases due to decreased bone density. As mentioned before, annual mortality rate of osteoporosis is higher than compared to cancers. Therefore, the use of data mining algorithms for the timely detection of this disease is very important. In this study, two different algorithms were used for this purpose and the results were compared to find the best prediction model in terms of accuracy and precision. The results showed that the random forest algorithm with 73.4% accuracy and 68.07% precision had a better performance in predicting osteoporosis in active middle-aged men.

As mentioned earlier, osteoporosis is treatable but is often overlooked. By combining predictive data mining algorithms in clinical practice, patients can benefit from early diagnosis and intervention. In the present study, a larger number of inputs were used that were easily measurable and studies have reported a close relationship between them and osteoporosis. It is obvious that the inclusion of these inputs provide more information and contribute to the better performance of algorithms. Health care providers collect a lot of data but this data is not used properly. By finding the hidden patterns and relationships in this data, they can be used properly to improve the quality of diagnostic and treatment services.

Ethical Considerations

Compliance with ethical guidelines

In the implementation of the research, ethical considerations were considered in accordance with the instructions of the ethics committee of Allameh Tabataba'i University and the code of ethics was received under the number IR.ATU.REC.1399.038.

Funding

This article is taken from Leila Fasihi's thesis under the guidance of Bakhtiar Zorgian and the advice of Rasul Islami in the Department of Exercise Physiology, Faculty of Physical Education and Sport Sciences, Allameh Tabataba'i University, Tehran.

Authors' contributions

All authors contributed equally in preparing all parts of the research.

Conflict of interest

The authors declared no conflict of interest.

Acknowledgments

The authors thank Allameh Tabataba'i University and the subjects who participate in the study.

مقاله پژوهشی

مقایسه دو الگوریتم درخت تصمیم‌گیری ساده و جنگل تصادفی برای پیش‌بینی بیماری پوکی استخوان در مردان میانسال فعال

*لیلا فصیحی[1]، بختیار ترتیبیان[1]، رسول اسلامی[1]

۱. گروه فیزیولوژی ورزشی، دانشکده تربیت‌بدنی و علوم ورزشی، دانشگاه علامه طباطبائی، تهران، ایران.

Use your device to scan and read the article online

Citation Fasihi L, Tartibian B, Eslami R. [Comparing the Effectiveness of Two Simple Decision Tree and Random Forest Algorithms in Predicting Osteoporosis in Active Middle-Aged Men (Persian)]. Scientific Journal of Rehabilitation Medicine. 2022; 11(2):264-275. https://dx.doi.org/10.32598/SJRM.11.2.10

https://dx.doi.org/10.32598/SJRM.11.2.10

CC BY

چکیده

مقدمه و اهداف پوکی استخوان، بیماری خاموش نامیده می‌شود، زیرا شکستگی اولین نشانه آشکارکننده این بیماری است. قبل از وقوع شکستگی، تشخیص زودهنگام این بیماری مسئله مهمی است. استفاده از الگوریتم‌های داده‌کاوی، دانشی است که ما را در رسیدن به اطلاعات نهفته در داده‌ها یاری می‌کند. هدف اصلی این مطالعه مقایسه دو الگوریتم درخت تصمیم‌گیری ساده و جنگل تصادفی برای پیش‌بینی بیماری پوکی استخوان در مردان میانسال فعال بود.

مواد و روش‌ها تعداد ۲۵۶ نفر مرد میانسال که طی سال‌های ۱۳۹۶ تا ۱۳۹۹ در بیمارستان آیت‌الله کاشانی تهران پرونده پزشکی داشتند و با عنوان مشکوک به بیماری پوکی استخوان به این بیمارستان مراجعه کرده بودند، در این مطالعه شرکت کردند. از نرم‌افزار متلب نسخه ۲۰۲۰ برای تجزیه‌وتحلیل داده‌ها استفاده شد. ارزیابی براساس ماتریس درهم‌ریختگی و معیارهای دقت و صحت انجام شد.

یافته‌ها ابتدا، طبق نظر پزشکان و نتایج مقالات علمی از ۱۰۳ ویژگی مربوط به اطلاعات شخصی، سبک زندگی و اطلاعات بیماری، ۱۱ ویژگی به‌عنوان ورودی الگوریتم‌ها انتخاب شدند. الگوریتم جنگل تصادفی با دقت ۷۳/۴ درصد و صحت ۶۸/۰۷ درصد عملکرد بهتری داشت.

نتیجه‌گیری در این مطالعه با توجه به داده‌های در دسترس، الگوریتم‌هایی توسعه داده شد که می‌تواند درزمینه پیشگیری بیماری پوکی استخوان کارآمد باشد. استفاده از این الگوریتم‌ها می‌تواند برای شروع درمان و تشخیص سریع این بیماری مؤثر باشد و از وقوع شکستگی و عوارض جبران‌ناپذیر آن جلوگیری کند.

کلیدواژه‌ها پوکی استخوان، پیش‌بینی، مردان، میانسال، فعال

تاریخ دریافت: ۲۵ اسفند ۱۳۹۹

تاریخ پذیرش: ۳۰ فروردین ۱۴۰۰

تاریخ انتشار: ۰۱ خرداد ۱۴۰۱

* **نویسنده مسئول:**

لیلا فصیحی

نشانی: تهران، دانشگاه علامه طباطبائی، دانشکده تربیت‌بدنی و علوم ورزشی، گروه فیزیولوژی ورزشی.

تلفن: ۲۱۸۳۶۰۴ (۹۱۸) ۹۸+

رایانامه: l_fasihi@atu.ac.ir

مقدمه

پوکی استخوان[1] نوعی بیماری است که در آن تراکم استخوان و کیفیت ریزساختار استخوان کاهش می‌یابد و به افزایش خطر شکستگی منجر می‌شود [۱]. این بیماری استخوان‌ها را نازک و مستعد شکستگی می‌کند و شایع‌ترین بیماری متابولیک استخوان است [۲]. این بیماری علائم زیادی ندارد و فرد مبتلا تنها زمانی به بیماری خود پی می‌برد که در اثر یک عطسه، سرفه یا زمین خوردن ساده تعدادی از استخوان‌هایش دچار شکستگی شوند [۳].

سازمان بهداشت جهانی[2] از پوکی استخوان به‌عنوان یکی از چهار بیماری شایع جهان نام می‌برد [۴]. امروزه پوکی استخوان یک تهدید بزرگ در جهان محسوب می‌شود و مرگ‌ومیر سالیانه آن بیش از انواع سرطان‌هاست [۵]. حدود ۲۰ درصد از بیمارانی که دچار شکستگی می‌شوند در مراکز پرستاری تخصصی نیاز به مراقبت طولانی‌مدت دارند. ۶۰ درصد بیماران هرگز استقلال عمل قبل از وقوع شکستگی را دوباره به‌دست نمی‌آورند [۶].

در مطالعه‌ای که وزارت بهداشت ایران انجام داده است، در ایران ۴۴ درصد مردان و ۴۷ درصد زنان بالای ۵۰ سال دچار کمبود تراکم استخوان و نیز ۴/۶ درصد افراد ۲۰ تا ۷۰ سال به پوکی استخوان در ستون فقرات مبتلا هستند [۷]. همچنین ۱ نفر از هر ۴ زن ایرانی بالای ۵۰ سال به پوکی استخوان مبتلاست. این در حالی است که در دنیا از هر ۳ زن، ۱ زن و از هر ۵ مرد ۱ مرد بالای ۵۰ سال دچار پوکی استخوان هستند [۸].

روش بررسی تراکم استخوان با استفاده از دستگاه دگزا[3] است. در این روش برای تشخیص افراد بیمار، مقدار تراکم استخوان آن‌ها با میانگین افراد جوان همجنس مقایسه می‌شود که به آن امتیاز تی[4] می‌گویند [۹]. بنابر راهنمایی‌های سازمان بهداشت جهانی، افرادی که امتیاز تی آن‌ها کمتر از ۲/۵- است، استئوپروز هستند و فاصله بین ۲/۵ تا ۱ انحراف معیار کمتر استئوپنی و افرادی که امتیاز تی آن‌ها بیشتر از ۱- باشد، در گروه افراد طبیعی قرار می‌گیرند [۱۰].

عوامل مختلفی در ایجاد پوکی استخوان دخالت دارند که بعضی مانند سن، جنس، نژاد، سابقه فامیلی، یائسگی غیرقابل کنترل و برخی مانند ورزش نکردن، کمبود کلسیم و ویتامین D در رژیم غذایی، قهوه، نمک، سیگار و الکل قابل کنترل هستند [۱۱]. مهم‌ترین عوامل مؤثر بر تراکم استخوان، عوامل ژنتیکی، نژادی و نوع و سبک خاص زندگی کنونی ازجمله کم‌تحرکی و بی‌تحرکی هستند [۱۲].

مطالعات مختلف اثر فعالیت بدنی را در پیشگیری از تحلیل بافت استخوان متذکر شدند و همچنین با توجه به اینکه ۴۰ تا ۴۴ درصد تراکم استخوانی یک فرد بالغ طی دوران نوجوانی به‌دست می‌آید، تأثیر فعالیت بدنی را به‌خصوص در دوران رشد و بلوغ روی تراکم استخوانی‌شان دادند [۱۳]. فعالیت ورزشی نیز به‌عنوان یک عامل نگه‌دارنده و محرک تشکیل استخوان است که ازطریق تجمع مواد معدنی، تقویت عضلات و بهبود تعادل فرد به کاهش ریسک شکستگی‌های استخوان منجر می‌شود [۱۴]. تأثیر فعالیت بدنی و فشار مکانیکی وارده ناشی از آن بر افزایش تراکم استخوانی به اثبات رسیده است، فشارهای مکانیکی واردشده بر استخوان ازطریق تاندون‌ها و عضلات اثر مستقیمی بر تشکیل استخوان و تغییر شکل آن دارد [۱۵]. فعالیت‌های ورزشی به دو روش سبب انتقال نیرو به استخوان می‌شود: کشش عضله و نیروی جاذبه. این نیروها می‌توانند سبب افزایش تراکم استخوان شوند. چنانچه افرادی که زندگی فعال را دنبال می‌کنند، نسبت به افراد غیرفعال هم‌سن‌وسال خود به نحو چشم‌گیری جرم استخوانی بیشتری دارند و این بهره تا دهه هفتم و حتی هشتم زندگی‌شان حفظ می‌شود [۱۶].

حجم عظیمی از داده‌های خام، ما را در دنیای امروزی احاطه کرده‌است، درحالی‌که نیازمند دانش خاصی هستیم که این حجم عظیم از داده بتواند ما را در جهت بهبود زندگی یاری دهد. علمی که ما را در رسیدن به دانش نهفته در داده‌ها یاری می‌کند، داده‌کاوی نامیده می‌شود [۱۷]. داده‌کاوی یک الگوی جدید برای تجزیه‌وتحلیل داده‌ها از بیماران و دستیابی به الگوهای مفید و علمی است [۱۸]. داده‌کاوی به‌خاطر داشتن مزایایی مثل قدرت پردازش حجم عظیم داده‌ها و کاهش زمان تشخیص درزمینه‌های مختلف ازجمله پزشکی و ورزش، کاربردهای زیادی دارد و به آنان در معنا دادن به داده‌های پیچیده کمک می‌کند. همچنین قدرت داده‌کاوی برای تشخیص رضایت‌بخش بیماری‌های گوناگون ثابت شده است [۱۹].

الگوریتم‌های داده‌کاوی مانند ماشین بردار پشتیبان، درخت تصمیم‌گیری و جنگل تصادفی با موفقیت در زمینه‌های مختلف ازجمله پزشکی و ورزش استفاده شدند [۲۰]. این الگوریتم‌ها قادر بودند تا راه‌حل معقول و مناسبی را در سیستم تشخیص بیماری‌هایی ازقبیل پوکی استخوان، بیماری‌های قلبی، سرطان سینه و دیابت ارائه دهند [۲۱]. کارشناسان اخیراً «بحران در درمان پوکی استخوان» را برجسته کردند، زیرا در سال‌های اخیر تجویز و پای‌بندی به رژیم‌های دارویی کاهش یافت که ممکن است به‌دلیل افزایش آگاهی از عوارض جانبی شدید (درعین‌حال نادر)، مصرف داروی اصلی درمان آن (بیس فسفونات‌ها) باشد [۲۲]. در حالی‌که آموزش برای پزشکان و بیماران به احتمال زیاد نکته کلیدی و مهم در بهبود نسخه و میزان انطباق داروهای پوکی استخوان است، اما این مشکلات اهمیت شناسایی روش‌های درمانی غیردارویی برای پیشگیری و درمان پوکی استخوان را برجسته می‌کند [۲۳]. برای سنجش تراکم استخوان، تشخیص

1. Osteoporosis
2. World Health Organization (WHO)
3. Dual Energy X-Ray Absorptiometry (DEXA)
4. T-Score

زودرس بیماری و همچنین کاهش هزینه‌های ناشی از عوارض آن انجمن‌های پزشکی مختلف، شناسایی براساس ریسک فاکتورها را توصیه می‌کنند. با توجه به ناکارآمدی این توصیه‌ها در طبابت بالینی به ابزارهای مختلف برای پیش‌بینی میزان پوکی استخوان و رتبه‌بندی آن در مردان و زنان نیاز است. الگوریتم‌های داده‌کاوی ابزاری ساده و کارآمد برای پیش‌بینی خطر پوکی استخوان است [۲۴].

استفاده از الگوریتم‌های داده‌کاوی باعث می‌شود اطلاعات جدید و روابط تعبیه‌شده در مجموعه‌های داده‌های بزرگ و پیچیده با استنتاج روابط جدید و یادگیری الگوها مشهود باشد [۲۵]. به کمک این الگوریتم‌ها می‌توان بدون نیاز به روش‌های تشخیصی، احتمال ابتلا به این بیماری را تشخیص داد و عوامل مؤثر در این بیماری را شناسایی کرد. بنابراین هدف اصلی این مطالعه، مقایسه دو الگوریتم درخت تصمیم‌گیری ساده و جنگل تصادفی برای پیش‌بینی بیماری پوکی استخوان در مردان میانسال فعال بود.

مواد و روش‌ها

اطلاعات بالینی استفاده‌شده در این مطالعه مربوط به بیماران مراجعه‌کننده به بیمارستان بود. از این افراد، تعداد ۲۵۶ مرد فعال انتخاب شدند که ۸۹ نفر سالم، ۱۲۶ نفر استئوپنی و ۴۱ نفر استئوپروز بودند و طی سال‌های ۱۳۹۶ تا ۱۳۹۹ با عنوان مشکوک به بیماری پوکی استخوان به بیمارستان آیت‌الله کاشانی تهران مراجعه کرده بودند و در فایل‌های بایگانی رایانه بیمارستان پرونده حاوی اطلاعات آزمایشگاهی داشتند. پرونده بیماران و اطلاعات آزمایشگاهی آن‌ها حاوی ۱۰۳ ویژگی ثبت‌شده مربوط به پوکی استخوان، در رابطه با اطلاعات شخصی، سبک زندگی و اطلاعات بیماری بود. برای کاهش اندازه ۱۰۳ ویژگی اولیه در داده‌ها، طبق نظر ۲ پزشک متخصص درزمینه غدد درون‌ریز و روماتولوژی و با کمک نتایج مقالات علمی در زمینه پوکی استخوان [۲۶-۲۸] نهایتاً ۱۱ ویژگی به‌عنوان ورودی الگوریتم‌ها انتخاب شد.

ویژگی‌های انتخاب‌شده شامل سن، قد، وزن، شاخص توده بدنی، فعالیت بدنی، استعمال دخانیات، دریافت ویتامین D و کلسیم، سابقه شکستگی، سابقه پوکی استخوان در بستگان درجه ۱، انحنای ستون فقرات و کمر درد بود. در این مطالعه، منظور از مردان فعال، افرادی بودند که به‌طور منظم حداقل ۱ سال، هفته‌ای ۳ جلسه فعالیت بدنی داشتند. پس از پر کردن فرم رضایت‌نامه از اطلاعات مربوط به آزمایش سنجش تراکم مواد معدنی استخوان[5] آن‌ها استفاده شد.

معیارهای ورود شامل جنسیت مرد سن بین ۵۰ تا ۶۵ ساله، دارای سوابق پزشکی و آزمایشات بالینی در بیمارستان، امتیاز تی بین ۲ تا ۲- شاخص توده بدنی، بین ۱۸ تا ۴۰ و در دسترس بودن ازطریق تلفن یا اینترنت بودند. معیارهای خروج شامل تحت درمان پوکی استخوان یا هورمون استروژن، سابقه مصرف داروهای هورمونی و داشتن بیماری مزمن بود. معیار تفکیک افراد با توجه به امتیاز تی (T-Score) آزمودنی‌ها در نواحی گردن استخوان ران بود. براساس استاندارد سازمان بهداشت جهانی امتیاز تی کمتر از ۲/۵- نشان‌دهنده ابتلای فرد به استئوپروز یا پوکی استخوان است، امتیاز تی بین ۱- تا ۲/۵- فرد دچار استئوپنی یا در حال فرسایش استخوانی و امتیاز تی بیشتر از ۱- فرد سالم است [۲۹].

برای پیش‌بینی و دسته‌بندی افراد سالم، استئوپنی و استئوپروز از ۲ الگوریتم داده‌کاوی، درخت تصمیم‌گیری ساده[6] و جنگل تصادفی[7] استفاده شد. در الگوریتم درخت تصمیم‌گیری سعی می‌شود از داده‌های آموزشی برای تصمیم‌گیری و تعیین کلاس داده‌های تست استفاده شود و به‌صورت ۱ درخت پیاده‌سازی و دسته‌بندی شوند، به‌طوری‌که در زمان ارزیابی کردن و با دیدن داده جدید بتوان کلاس مربوط به آن را پیش‌بینی کرد [۳۰]. در الگوریتم جنگل تصادفی چندین درخت تصمیم‌گیری ساخته شده و آن‌ها را با یکدیگر ادغام می‌کند تا پیش‌بینی‌های صحیح‌تر و پایدارتری حاصل شود (تصویر شماره ۱) [۳۱].

برای استفاده بهینه از داده‌ها باید آن‌ها را به شکلی تغییر داد که برای الگوریتم‌های داده‌کاوی مناسب باشند [۳۲]. برای پرسش‌هایی که پاسخ بله و خیر داشتند از عدد صفر و ۱ استفاده شد. صفر به معنای پاسخ خیر و عدد ۱ به معنای پاسخ بله درنظر گرفته شد. در این مرحله، مجموعه داده‌ها در محیط اکسل به نرم‌افزار متلب[8] نسخه ۲۰۲۰ انتقال داده و تجزیه‌وتحلیل شد. معیارهای دقت و صحت براساس روش ارزیابی داده‌ها (جدول شماره ۱) نشان می‌دهد دقت معادل «چه میزان از نمونه‌های انتخابی درست هستند» و صحت معادل «چه میزان از نمونه‌های صحیح موجود، درست انتخاب شدند» [۳۳].

الگوریتم‌ها براساس دقت[9] و صحت[10] با یکدیگر مقایسه شدند. دقت الگوریتم ارزش آن را در پیش‌بینی نشان می‌دهد از تعداد پیش‌بینی‌های صحیح، تقسیم بر تعداد کل پیش‌بینی‌ها به‌دست می‌آید (فرمول شماره ۱). صحت الگوریتم نشان‌دهنده قدرت تفکیک آن برای جدا کردن افراد بیمار و سالم از یکدیگر است و از تقسیم تعداد پیش‌بینی بر تعداد پیش‌بینی‌های هر ردیف به‌دست می‌آید (فرمول شماره ۲).

۱. دقت

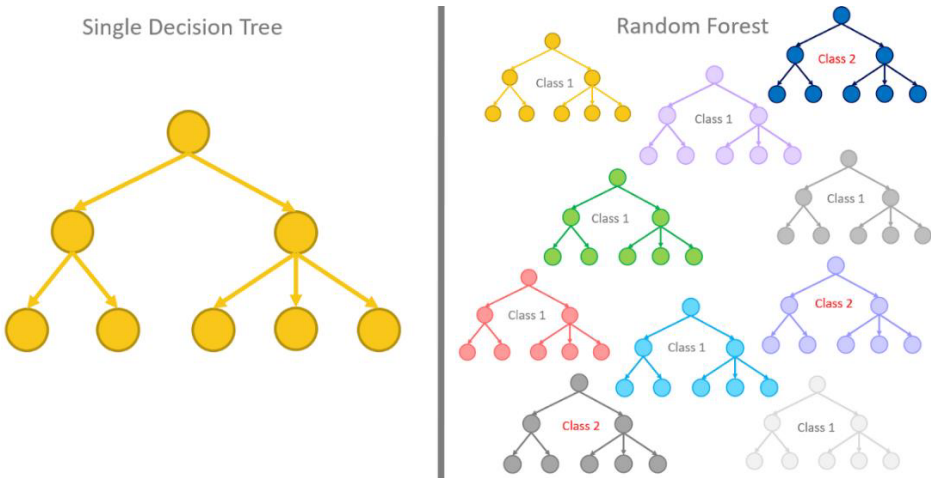
$$Accuracy=\frac{TP+TN}{TP+FP+TN+FN}$$

۲. صحت

$$Precision=\frac{TP}{TP+FP}$$

5. Dual-Energy X-Ray Absorptiometry (DEXA)

6. Simple Decision Tree
7. Random Forest
7. Matlab
9. Accuracy
10. Precision



تصویر ۱. تفاوت بین الگوریتم درخت تصمیم و جنگل تصادفی

طب توانبخشی

یافته‌ها

اطلاعات توصیفی آزمودنی‌ها و متغیرهای شناسایی‌شده براساس نتایج مقالات علمی و نظرسنجی از پزشکان در جدول شماره ۲ نشان داده شدند. ۲۵ درصد داده‌ها برای تست و ۷۵ درصد برای آموزش الگوریتم درنظر گرفته شدند. از دو الگوریتم درخت تصمیم‌گیری ساده و جنگل تصادفی برای پیش‌بینی پوکی استخوان استفاده شد و نتایج با هم مقایسه شدند. تصاویر شماره ۲ و ۳ نتایج حاصل از ماتریس درهم‌ریختگی دو الگوریتم نشان داده شده است. با توجه به نتایج تصاویر شماره ۲ و ۳ مشاهده می‌شود الگوریتم جنگل تصادفی با دقت ۷۳/۴ درصد و صحت ۶۸/۰۷ درصد عملکرد بهتری دارد.

بحث

محتوای مواد معدنی استخوان در دوران کودکی افزایش می‌یابد و در دوران بلوغ به حداکثر میزان خود می‌رسد. بعد از دهه سوم زندگی با افزایش سن کاهش کلی استخوان شروع می‌شود و قدرت استخوان به‌علت کاهش تراکم استخوان کم می‌شود [۳]. امروزه مرگ‌ومیر سالیانه ناشی از پوکی استخوان بیش از انواع سرطان‌هاست. بنابراین استفاده از الگوریتم‌های داده‌کاوی در تشخیص به‌موقع این بیماری اهمیت زیادی دارد [۳۴].

هدف از این تحقیق، مقایسه دو الگوریتم درخت تصمیم‌گیری ساده و جنگل تصادفی برای پیش‌بینی پوکی استخوان در مردان میانسال فعال بود. نتایج نشان داد با استفاده از ۱۱ ویژگی بالینی، الگوریتم جنگل تصادفی با دقت ۷۳/۴ درصد و صحت ۶۸/۰۷

جدول ۱. روش ارزیابی داده‌ها

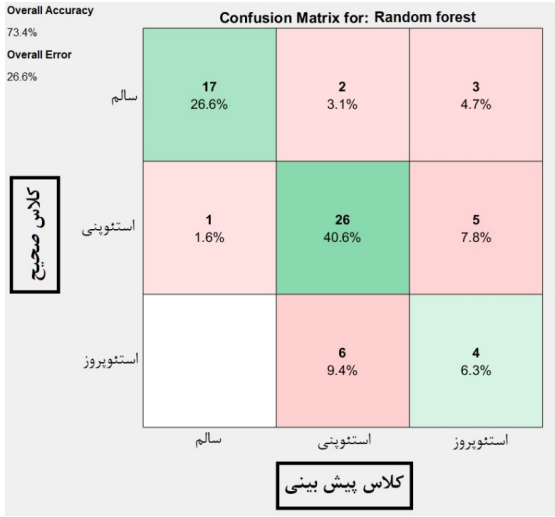
مقادیر پیش بینی شده		مقادیر صحیح	
		مثبت	منفی
مقادیر پیش‌بینی‌شده	مثبت	درست مثبت	نادرست مثبت
	منفی	نادرست منفی	درست منفی

طب توانبخشی

جدول۲. اطلاعات فردی آزمودنی‌ها

شاخص‌ها	میانگین±انحراف معیار
سن(سال)	۵۷/۳۰±۷/۶۷
قد(سانتی‌متر)	۱۷۱/۸۷±۸/۳۲
وزن(کیلوگرم)	۷۷/۷۹±۱۲/۶۲
شاخص توده بدنی(کیلوگرم بر متر مربع)	۲۶/۴۷±۴/۱۷

طب توانبخشی



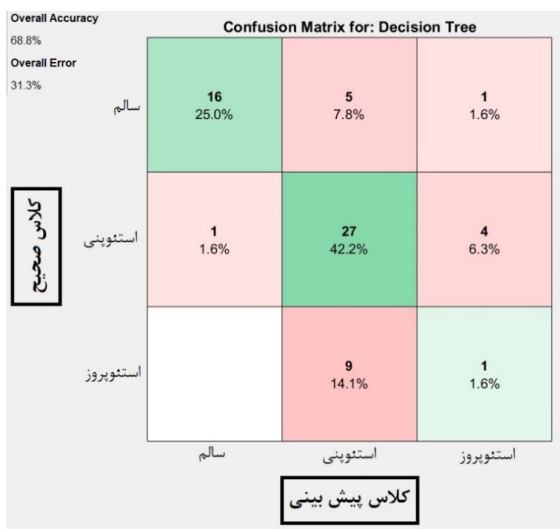
تصویر ۲. ماتریس درهم‌ریختگی الگوریتم جنگل تصادفی

طب توانبخشی

درصد در پیش‌بینی پوکی استخوان مردان میانسال عملکرد بهتری داشت. مطالعات گوناگون ویژگی‌های متفاوت و روش‌های داده‌کاوی مختلفی را برای پیش‌بینی پوکی استخوان استفاده کردند که تعدادی از آن‌ها با نتایج مطالعه فوق همسو بود.

ایلیو و همکاران، طی مطالعه‌ای در کشور یونان از الگوریتم‌های مختلف داده‌کاوی برای پیش‌بینی پوکی استخوان استفاده کردند. شاخص‌های استفاده‌شده در پژوهش آن‌ها شامل سن، جنسیت، وزن و قد بود. هدف آن‌ها استفاده از شاخص‌های کمتر برای پیش‌بینی پوکی استخوان با استفاده از الگوریتم‌های مختلف بود. آن‌ها در مطالعه خود با استفاده از اطلاعات ۱۰۸۳ نفر بیمار و ۲۳۴۳ نفر سالم به این نتیجه رسیدند که شبکه عصبی پرسپترون چندلایه با استفاده از تعداد اندکی از عوامل بالینی می‌تواند به پزشکان برای تشخیص پوکی استخوان کمک کند [۳۵].

در تحقیقی دیگر، ژوان و همکاران با استفاده از الگوریتم‌های طبقه‌بندی برای پیش‌بینی پوکی استخوان زنان تایوان از متغیرهای مختلف ازقبیل، اطلاعات جمعیت‌شناختی، قد، وزن، شاخص توده بدنی، اطلاعات آزمایشگاهی، سوابق پزشکی، اطلاعات مربوط به بیماری قلبی و پوکی استخوان برای پیش‌بینی پوکی استخوان استفاده کردند و برای الگوریتم‌های طبقه‌بندی استفاده شد. در مطالعه آن‌ها همانند مطالعه حاضر از متغیرهای موردنظر متخصصان استفاده شد. آن‌ها سعی کردند ازطریق الگوریتم، طبقه‌بندی تلفیقی را برای توسعه مدلی به‌کار گیرند که براساس آن بتوان پوکی استخوان را پیش‌بینی کرد. درنهایت، الگوریتم درخت تقویت‌شده بهترین نتایج را برای پیش‌بینی ارائه کرد که با نتایج تحقیق حاضر همسو بود [۳۶].



تصویر ۳. ماتریس درهم‌ریختگی الگوریتم درخت تصمیم‌گیری ساده

طب توانبخشی

علیزاده و همکاران در مطالعه خود بر نتایج سنجش تراکم استخوان به مقایسه الگوریتم‌های داده‌کاوی پرداختند. در این مطالعه، براساس سبک زندگی، بیماری‌ها و نتایج تراکم استخوانی از ۶۰ متغیر با ۴ بخش اطلاعات فردی، سبک زندگی، اطلاعات بیماری و نتایج آزمایش دگزا استفاده شد. در این مطالعه، مهم‌ترین عوامل بالینی برای استفاده در الگوریتم‌ها شامل سن یائسگی، قرارگیری روزانه در معرض نور آفتاب، ابتلا به قوز یا گودی کمر، درد زانو، مصرف دوغ در هفته، استفاده از دندان مصنوعی، مصرف آب، میزان مصرف چای، ورزش و کوتاه شدن قد بود. آن‌ها پس از مقایسه روش‌های مختلف داده‌کاوی برای بررسی پوکی استخوان نشان دادند شبکه عصبی مصنوعی قدرتمندتر از الگوریتم‌های درخت تصمیم است که با مطالعه فوق ناهمسو بود [۳۷]. شاید علت همسو نبودن، مربوط به نوع و تعداد شاخص‌های درنظر گرفته‌شده باشد. آن‌ها در مطالعه خود از ۱۰ شاخص استفاده کرده بودند، درحالی‌که در مطالعه حاضر از ۱۹ شاخص استفاده شد.

همچنین مطالعه غفوری و همکاران با مطالعه فوق ناهمسو بود. آن‌ها در مطالعه خود برای بررسی میزان پوکی استخوان و وقوع شکستگی در ایران از ابزار فرکس استفاده کرده بودند و به این نتیجه رسیدند که روش فرکس برای جمعیت ایران باید تغییر کند، زیرا بسیاری از عوامل استفاده‌شده در این ابزار با وقوع شکستگی در بین جمعیت ایران مرتبط نیستند. این مطالعه عواملی مثل بی‌تحرکی برای بیش از ۹۰ روز را از عوامل مهم مرتبط با وقوع شکستگی معرفی می‌کند. همچنین دیابت به‌عنوان یکی از عواملی معرفی شد که خطر شکستگی را افزایش می‌دهد. به‌علاوه، مصرف داروهای کورتیکواستروئید نیز عامل مهمی شناخته شد. درنهایت، این مطالعه پیشنهاد داد عوامل جدیدی باید در بررسی پوکی استخوان درنظر گرفته شوند [۳۸].

احتمالاً علت عدم همسویی با مطالعه فوق می‌تواند به‌علت تفاوت در ابزار داده‌کاوی (روش فرکس) استفاده شده و همچنین تفاوت در شاخص‌های استفاده‌شده در تحقیق آن‌ها ازقبیل دیابت و مصرف داروهای کورتیکواستروئید باشد. چنانکه قبلاً ذکر شد، پوکی استخوان قابل درمان است، اما اغلب نادیده گرفته می‌شود. با ترکیب الگوریتم‌های پیش‌بینی در عمل بالینی، بیماران می‌توانند از تشخیص و آزمایش زودهنگام بهره ببرند.

در مطالعه حاضر، از تعداد بالای شاخص‌هایی استفاده شده است که هم به سادگی قابل اندازه‌گیری هستند و هم اینکه در مطالعات اخیر ارتباط نزدیک این شاخص‌ها با پوکی استخوان گزارش شده‌است. مشهود است که گنجاندن این شاخص‌ها اطلاعات بیشتری ارائه می‌کند و به عملکرد بهتر در الگوریتم‌های پیش‌بینی کمک می‌کند.

یکی از محدودیت‌های تحقیق حاضر، کم بودن پرونده پزشکی آزمودنی‌های ۵۰ تا ۶۵ ساله دسترس، به‌ویژه افراد فعال بود. با توجه به اهمیت رژیم غذایی در تراکم استخوان افراد، عدم امکان تعیین نقش رژیم غذایی استفاده‌شده آزمودنی‌ها، محدودیت دیگر این مطالعه بود. هرچه تعداد آزمودنی‌ها بیشتر باشد، نتایج دقیق‌تر و کامل‌تری از پیش‌بینی بیماری پوکی استخوان به‌دست می‌آید و می‌تواند در شناسایی افراد در معرض پوکی استخوان و ارتقای کیفیت زندگی و پیشگیری از شکستگی‌های ناشی از آن استفاده شود. ازاین‌رو، پیشنهاد می‌شود در مطالعات آینده از تعداد آزمودنی‌های بیشتری استفاده شود.

نتیجه‌گیری

در این مطالعه، دو الگوریتم مختلف بررسی و نتایج برای مقایسه بهترین مدل پیش‌بینی از نظر دقت و صحت مقایسه شدند. با توجه به داده‌های در دسترس، احتمالاً هر دو الگوریتم می‌توانند درزمینه پیشگیری بیماری پوکی استخوان کارآمد باشند. استفاده از این الگوریتم‌ها می‌تواند برای شروع درمان و تشخیص سریع این بیماری مؤثر باشد و از وقوع شکستگی و عوارض جبران‌ناپذیر آن جلوگیری کند. در آخر، توجه به این نکته مهم است که سازمان‌های مراقبت‌های بهداشتی همیشه اطلاعات زیادی جمع‌آوری می‌کنند، درحالی‌که از این اطلاعات و داده‌ها به‌درستی استفاده نمی‌شود. این مطالعه نشان می‌دهد با کشف الگوها و روابط پنهان در این داده‌ها می‌توان به‌درستی از این داده‌ها برای بهبود کیفیت خدمات تشخیصی و درمانی استفاده کرد.

ملاحظات اخلاقی

پیروی از اصول اخلاق پژوهش

در اجرای پژوهش، ملاحظات اخلاقی مطابق با دستورالعمل کمیته اخلاق دانشگاه علامه طباطبائی درنظر گرفته شده و کد اخلاق به شماره IR.ATU.REC.1399.038 دریافت شده است.

حامی مالی

این مقاله برگرفته از پایان‌نامه لیلا فصیحی با راهنمایی بختیار ترتیبیان و مشاوره رسول اسلامی در گروه فیزیولوژی ورزشی دانشگاه علامه طباطبائی است.

مشارکت نویسندگان

تمام نویسندگان در آماده‌سازی این مقاله مشارکت یکسان داشته‌اند.

تعارض منافع

بنابر اظهار نویسندگان، این مقاله تعارض منافع ندارد.

تشکر و قدردانی

از تمام شرکت‌کنندگان در این پژوهش و افرادی که ما را در انجام پژوهش حاضر یاری رساندند، قدردانی می‌شود.

References

[1] Ominsky MS, Boyce RW, Li X, Ke HZ. Effects of sclerostin antibodies in animal models of osteoporosis. Bone. 2017; 96:63-75. [DOI:10.1016/j.bone.2016.10.019] [PMID] [PMCID]

[2] Larijani B, Hossein-Nezhad A, Mojtahedi A, Pajouhi M, Bastanhagh MH, Soltani A, et al. Normative data of bone mineral density in healthy population of Tehran, Iran: A cross sectional study. BMC Musculoskeletal Disorders. 2005; 6(1):1-6. [DOI:10.1186/1471-2474-6-38] [PMID] [PMCID]

[3] Bouxsein ML, Eastell R, Lui LY, Wu LA, de Papp AE, Grauer A, et al. Change in bone density and reduction in fracture risk: A meta-regression of published trials. Journal of Bone and Mineral Research. 2019; 34(4):632-42. [DOI:10.1002/jbmr.3641] [PMID]

[4] Sözen T, Özışık L, Başaran NÇ. An overview and management of osteoporosis. European Journal of Rheumatology. 2017; 4(1):46-56. [DOI:10.5152/eurjrheum.2016.048] [PMID] [PMCID]

[5] Baidoo PK, Odei JB, Ansu V, Segbefia M, Holdbrook-Smith H. Predictors of hip fracture mortality in Ghana: A single-center prospective study. Archives of Osteoporosis. 2021; 16(1):35. [DOI:10.1007/s11657-021-00883-z] [PMID]

[6] Golob AL, Laya MB. Osteoporosis: Screening, prevention, and management. Medical Clinics of North America. 2015; 99(3):587-606. [DOI:10.1016/j.mcna.2015.01.010] [PMID]

[7] Keramat A, Patwardhan B, Larijani B, Chopra A, Mithal A, Chakravarty D, et al. The assessment of osteoporosis risk factors in Iranian women compared with Indian women. BMC Musculoskeletal Disorders. 2008; 9:28. [DOI:10.1186/1471-2474-9-28] [PMID] [PMCID]

[8] Fahimfar N, Noorali S, Yousefi S, Gharibzadeh S, Shafiee G, Panahi N, et al. Prevalence of osteoporosis among the elderly population of Iran. Archives of Osteoporosis. 2021; 16(1):16. [DOI:10.1007/s11657-020-00872-8] [PMID]

[9] Biedenweg K, Meischke H, Bohl A, Hammerback K, Williams B, Poe P, et al. Understanding older adults' motivators and barriers to participating in organized programs supporting exercise behaviors. The Journal of Primary Prevention. 2014; 35(1):1-11. [DOI:10.1007/s10935-013-0331-2] [PMID]

[10] Hong N, Park H, Rhee Y. Machine learning applications in endocrinology and metabolism research: An overview. Endocrinology and Metabolism. 2020; 35(1):71-84. [DOI:10.3803/EnM.2020.35.1.71] [PMID] [PMCID]

[11] Hirota T, Nara M, Ohguri M, Manago E, Hirota K. Effect of diet and lifestyle on bone mass in Asian young women. The American Journal of Clinical Nutrition. 1992; 55(6):1168-73. [DOI:10.1093/ajcn/55.6.1168] [PMID]

[12] Hughes D, Mikosch P, Belmatoug N, Carubbi F, Cox T, Goker-Alpan O, et al. Gaucher disease in bone: From pathophysiology to practice. Journal of Bone and Mineral Research. 2019; 34(6):996-1013. [DOI:10.1002/jbmr.3734] [PMID] [PMCID]

[13] Nuti R, Brandi ML, Checchia G, Di Munno O, Dominguez L, Falaschi P, et al. Guidelines for the management of osteoporosis and fragility fractures. Internal and Emergency Medicine. 2019; 14(1):85-102. [DOI:10.1007/s11739-018-1874-2] [PMID] [PMCID]

[14] Espinosa R, Clark P, Denova-Gutiérrez E, de los Ángeles Aguilera-Barreiro M, Flores M, Diez P, et al. Prevention of low bone mass to achieve high bone density in Mexico: Position of the Mexican association for bone and mineral metabolism. Archives of Osteoporosis. 2018; 13(1):1-10. [DOI:10.1007/s11657-018-0520-y] [PMID]

[15] Tong X, Chen X, Zhang S, Huang M, Shen X, Xu J, et al. The effect of exercise on the prevention of osteoporosis and bone angiogenesis. BioMed Research International. 2019; 2019:8171897. [DOI:10.1155/2019/8171897] [PMID]

[16] Daly RM, Gianoudis J, Kersh ME, Bailey CA, Ebeling PR, Krug R, et al. Effects of a 12-month supervised, community-based, multimodal exercise program followed by a 6-month research-to-practice transition on bone mineral density, trabecular microarchitecture, and physical function in older adults: A randomized controlled trial. Journal of Bone and Mineral Research. 2020; 35(3):419-29. [DOI:10.1002/jbmr.3865] [PMID]

[17] Cuaya-Simbro G, Perez-Sanpablo AI, Morales EF, Quiñones Uriostegui I, Nuñez-Carrera L. Comparing machine learning methods to improve fall risk detection in elderly with osteoporosis from balance data. Journal of Healthcare Engineering. 2021; 2021:8697805. [DOI:10.1155/2021/8697805] [PMID] [PMCID]

[18] Tartibian B, Fasihi L, Eslami R. Prediction of osteoporosis by K-NN algorithm and prescribing physical activity for elderly women. New Approaches in Sport Sciences. 2020; 2(4):87-100. [DOI:10.22054/NASS.2021.59130.1088]

[19] Viceconti M, Hunter P, Hose R. Big data, big knowledge: Big data for personalized healthcare. IEEE Journal of Biomedical and Health Informatics. 2015; 19(4):1209-15. [DOI:10.1109/JBHI.2015.2406883] [PMID]

[20] Ngiam KY, Khor W. Big data and machine learning algorithms for health-care delivery. The Lancet Oncology. 2019; 20(5):e262-73. [DOI:10.1016/S1470-2045(19)30149-4]

[21] Alonso SG, de la Torre Diez I, Rodrigues JJ, Hamrioui S, Lopez-Coronado M. A systematic review of techniques and sources of big data in the healthcare sector. Journal of Medical Systems. 2017; 41(11):183. [DOI:10.1007/s10916-017-0832-2] [PMID]

[22] Goldshtein I, Rouach V, Shamir-Stein N, Yu J, Chodick G. Role of side effects, physician involvement, and patient perception in non-adherence with oral bisphosphonates. Advances in Therapy. 2016; 33(8):1374-84. [DOI:10.1007/s12325-016-0360-3] [PMID]

[23] McMillan LB, Zengin A, Ebeling PR, Scott D. Prescribing physical activity for the prevention and treatment of osteoporosis in older adults. Healthcare (Basel). 2017; 5(4):85. [DOI:10.3390/healthcare5040085] [PMID] [PMCID]

[24] Bach M, Werner A, Żywiec J, Pluskiewicz W. The study of under-and over-sampling methods' utility in analysis of highly imbalanced data on osteoporosis. Information Sciences. 2017; 384:174-90. [DOI:10.1016/j.ins.2016.09.038]

[25] Wang Y, Wang L, Sun Y, Wu M, Ma Y, Yang L, et al. Prediction model for the risk of osteoporosis incorporating factors of disease history and living habits in physical examination of population in Chongqing, Southwest China: Based on artificial neural network. BMC Public Health. 2021; 21(1):991. [DOI:10.1186/s12889-021-11002-5] [PMID] [PMCID]

[26] Zanker J, Duque G. Osteoporosis in older persons: Old and new players. Journal of the American Geriatrics Society. 2019; 67(4):831-40. [DOI:10.1111/jgs.15716] [PMID]

[27] Xu N, Wang Y, Xu Y, Li L, Chen J, Mai X, et al. Effect of subclinical hyperthyroidism on osteoporosis: A meta-analysis of cohort studies. Endocrine. 2020; 69(1):39-48. [DOI:10.1007/s12020-020-02259-8] [PMID]

[28] Elkady AA, Kazem HH, Elgendy EA. Protective effect of vitamin D against rats' mandibular osteoporosis induced by corticosteroids and gamma rays. International Journal of Radiation Research. 2020; 18(1):125-31. [DOI:10.18869/acadpub.ijrr.18.1.125]

[29] Akkawi I, Zmerly H. Osteoporosis: Current concepts. Joints. 2018; 6(02):122-7. [DOI:10.1055/s-0038-1660790] [PMID] [PMCID]

[30] Bifet A, Zhang J, Fan W, He C, Zhang J, Qian J, et al. Extremely fast decision tree mining for evolving data streams. Paper presented in: The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017 Aug 13-17; Halifax: Canada. [DOI:10.1145/3097983.3098139]

[31] Alzubi J, Nayyar A, Kumar A. Machine learning from theory to algorithms: An overview. Paper presented in: Second National Conference on Computational Intelligence (NCCI 2018). 5 December 2018; Bangalore: India. [DOI:10.1088/1742-6596/1142/1/012012]

[32] Majumdar J, Naraseeyappa S, Ankalaki S. Analysis of agriculture data using data mining techniques: Application of big data. Journal of Big data. 2017; 4(1):1-5. [DOI:10.1186/s40537-017-0077-4]

[33] Bahrami B, Shirvani MH. Prediction and diagnosis of heart disease by data mining techniques. Journal of Multidisciplinary Engineering Science and Technology (JMEST). 2015; 2(2):164-8. [Link]

[34] Shirzadfar H, Gordoghli N. A comparative study of current methods and recent advances in the diagnosis and assessment of osteoporosis. Recent Research in Endocrinology and Metabolic Disorder. 2021; 3(1):37-9. [DOI:10.33702/rremd.2020.2.1.2]

[35] Iliou T, Anagnostopoulos C-N, Stephanakis IM, Anastassopoulos G. A novel data preprocessing method for boosting neural network performance: A case study in osteoporosis prediction. Information Sciences. 2017; 380:92-100. [DOI:10.1016/j.ins.2015.10.026]

[36] Juan Y-C, Chen C-M, Chen S-H. A classifier fusion approach to osteoporosis prediction for women in Taiwan. Journal of Industrial and Production Engineering. 2015; 32(6):360-8. [DOI:10.1080/21681015.2015.1064484]

[37] Alizadeh S, Sharifkhani M, Abbasi M, Ameri H. [Providing a model for predicting the risk of osteoporosis using decision tree algorithms (Persian)]. Journal of Mazandaran University of Medical Sciences. 2014; 24(116):110-8. [Link]

[38] Ghafoori S, Keshtkar A, Khashayar P, Ebrahimi M, Ramezani M, Mohammadi Z, et al. The risk of osteoporotic fractures and its associating risk factors according to the FRAX model in the Iranian patients: A follow-up cohort. Journal of Diabetes & Metabolic Disorders. 2014; 13(1):93. [DOI:10.1186/s40200-014-0093-2] [PMID] [PMCID]