scientific reports

Check for updates

OPEN

A new model for determining risk of male infertility from serum hormone levels, without semen analysis

Hideyuki Kobayashi✉, Masato Uetani, Fumito Yamabe, Yozo Mitsui, Koichi Nakajima & Koichi Nagao

We investigated a screening method using only serum hormone levels and AI (artificial intelligence) predictive analysis. Among 3662 patients, numbers for NOA (non-obstructive azoospermia), OA (obstructive azoospermia), cryptozoospermia, oligozoospermia and/or asthenozoospermia, normal, and ejaculation disorder were 448, 210, 46, 1619, 1333, and 6, respectively. "Normal" was defined as semen findings normal according to the WHO (World Health Organization) Manual for Human Semen Testing of 2021. We extracted age, LH (luteinizing hormone), FSH (follicle stimulating hormone), PRL (prolactin), testosterone, E2 (estradiol), and T (testosterone)/E2 from medical records. A total motility sperm count of 9.408 × 10^6 (1.4 ml × 16 × 10^6/ml × 42%) was defined as the lower limit of normal. The Prediction One-based AI model had an AUC (area under the curve) of 74.42%. For the AutoML Tables-based model, AUC ROC (receiver operating characteristic) was 74.2% and AUC PR (precision–recall) 77.2%. In a ranking of feature importance from 1st to 3rd, FSH came a clear 1st. T/E2 and LH ranked 2nd and 3rd for both Prediction One and AutoML Tables. Using data from 2021 and 2022 to verify the Prediction One-based AI model, the predicted and actual results for NOA were 100% matched in both years.

Keywords Artificial intelligence, Hormonal evaluation, Male infertility, Machine learning, Semen analysis

Infertility is a problem estimated to affect 72.4 million females and males worldwide. The WHO estimates that 9% of couples worldwide struggle with fertility problems and that male factors are involved in 50% of them[1,2].

Conventional semen analysis is the first step in the diagnosis of male infertility. The results of semen analysis reflect spermatogenesis in the testes, the patency of seminal ducts, and the glandular secretory activity[3]. The evaluation of semen parameters is currently based on the standards defined in the Laboratory Manual for the Examination and Processing of Human Semen created by WHO[4].

In addition to the semen analysis, serum hormonal levels are measured when investigating for male infertility. LH, FSH, total testosterone, E2, PRL, and T/E2 are measured in male infertility testing in the clinical setting.

Semen analysis and serum hormone levels indicate testicular function and the endocrine status of the hypothalamic-pituitary–testicular axis. Pulsatile secretion of GnRH stimulates FSH and LH secretion from the anterior pituitary. FSH stimulates Sertoli cells to induce spermatogenesis[5]. Sertoli cells secrete inhibin B and Leydig cells secrete testosterone. Testosterone is metabolized to E2 by aromatase. Inhibin B and E2 have negative feedback effects at the hypothalamic and pituitary levels. Both FSH and testosterone are required for spermatogenesis[6]. FSH is often elevated in spermatogenic dysfunction, but LH and testosterone secretion may be preserved[7,8].

Significant relationships between semen analysis results and serum hormone levels were reported some time ago[9]. A later study also found associations of FSH, LH and testosterone levels with semen analysis results[10].

However, since conventional sperm analysis methods involve complex, manual inspection with a microscope, they are labor intensive. In addition, many men are unwilling to be tested due to social stigma in certain parts of the world[11]. Although there are at-home sperm analysis kits for people to examine sperm condition by themselves, they are not a substitute for laboratory analysis[12]. Therefore, to address these issues, we believe that searching for a new method to determine the risk of male infertility is a valid research question.

Department of Urology, Toho University School of Medicine, 6-11-1, Omori-Nishi, Ota-Ku, Tokyo 143-8541, Japan. ✉email: hideyukk@med.toho-u.ac.jp

Since semen analysis results and serum hormone levels are correlated, we examined whether machine learning could be used to predict potential male infertility from serum hormone levels alone.

Machine learning is an applied technology for data that supports AI. A definition of machine learning is the field of study that gives computers the ability to learn without being explicitly programmed. Statistics and machine learning are quite different. Statistics was born out of statistical research, while machine learning was born out of computer science. They are very different in terms of data acquisition. Statistics is based on processing the data collected for a purpose. Machine learning does not necessarily involve data obtained with a purpose from the beginning. In other words, statistics focuses on collecting and analyzing data to confirm hypotheses and estimates, whereas machine learning starts from the data, searches for regularities and features in it, and applies them as AI. The advantage of machine learning is that it can deal with the huge amounts of big data.

If a successful screening system applying this became commercially available, the risk of male infertility could be determined without the need for semen analysis.

Results

The mean age of the 3662 patients who underwent sperm analysis and serum hormone level measurement for male infertility from 2011 to 2020 was 36.271 years old (95% CI (confidence interval)) 36.038–36.505). The means of LH, FSH, PRL, testosterone, E2, and T/E2 were 5.681 mIU/mL (95% CI 5.545–5.817), 8.845 mIU/mL (95% CI 8.535–9.155), 10.540 ng/mL (95% CI 9.865–11.214), 4.741 ng/mL (95% CI 4.672–4.810), 26.166 pg/mL (95% CI 25.802–26.530), and 19.917 (95% CI 19.544–20.290), respectively. The detailed data are shown in Table 1.

The means of sperm volume, concentration, motility, and total sperm motility count were 2.849 ml (95% CI 2.795–2.903), 41.830 X 10^6/ml (95% CI 28.858–54.801), 33.809% (95% CI 33.002–34.616), and 41.859×10^6 (95% CI 39.011–44.706), respectively. The detailed data are shown in Table 2.

Regarding male infertility, the 3662 patients were classified according to the results of semen analysis. Rates for NOA, OA, cryptozoospermia, oligozoospermia and/or asthenozoospermia, normal, and ejaculation disorder were 12.23% (n = 448), 5.73% (n = 210), 1.26% (n = 46), 44.21% (n = 1619), 36.40% (n = 1333), and 0.16% (n = 6), respectively. Table 3 shows the means for age, LH, FSH, PRL, testosterone, E2, T/E2, and sperm analyses (volume, concentration, motility, and total sperm motility count) for NOA, OA, cryptozoospermia, oligozoospermia or/and asthenozoospermia, normal, and ejaculation disorder.

Figure 1 shows the AI prediction model for risk of male infertility generated using Prediction One software. The ROC curve is a "guess curve". It is drawn by plotting sensitivity (true positive rate) on the vertical axis and specificity (false positive rate) on the horizontal axis. Generally, the larger the AUC, the better the machine learning performance. When determining normality or abnormality in a binary classification (0: normal, 1: abnormal) the machine learning model asks the question, "What is the probability of abnormality?" For example, when the threshold is 0.3, if the probability is greater than 0.3, this is determined as abnormal, and if it is less than 0.3, it is determined as normal. Increasing the threshold value increases "Precision", which indicates accuracy, but decreases "Recall", which indicates comprehensiveness. Precision and Recall are in a trade-off relationship.

In the accuracy evaluation of the AI prediction model, AUC was 74.42%. In a ranking of the contribution of variables from 1st to 7th, FSH came a clear 1st. The ranking order for 2nd and below was T/E2, LH, age, testosterone, E2 and PRL. The values for Accuracy, Precision and Recall, and the F-value were 63.39%, 56.61%, 82.53%, and 67.16%, respectively, when the threshold was 0.30. In addition, when the threshold was 0.49, the values for Accuracy, Precision and Recall, and the F-value were 69.67%, 76.19%, 48.19%, and 59.04%, respectively.

	n	Mean	95% confidence interval
Age (y) LH	3662	36.271	36.038–36.505
(mIU/mL)	3652	5.681	5.545–5.817
FSH (mIU/mL)	3653	8.845	8.535–9.155
PRL (ng/mL)	3652	10.54	9.865–11.214
Testosterone (ng/mL)	3655	4.741	4.672–4.810
E2 (pg/mL)	3645	26.166	25.802–26.530
T (ng/dL)/E2 (pg/mL)	3642	19.917	19.544–20.290

Table 1. Means of age and serum hormone levels (LH, FSH, PRL, testosterone, E2, and T/E2).

Total:3662	n	Mean	95% confidence interval
Sperm volume (mL)	3655	2.849	2.795–2.903
Concentration (X10^A6/mL)	3656	41.83	28.858–54.801
Motility (%)	3656	33.809	33.002–34.616
Total sperm motility count (X10A6/mL)	3662	41.859	39.011–44.706

Table 2. Means of sperm analysis (volume, concentration, motility, and total sperm motility count).

		n	Mean	95% confidence interval
Age (y)	NOA	448	35.433	34.802–36.064
	OA	210	35.81	34.846–36.773
	Cryptozoospermia	46	37.044	34.659–39.428
	Oligoasthenozoospermia	1619	37.473	37.11–37.836
	Normal	1333	35.139	34.776–35.502
	Ejaculation disorder	6	36.5	19.543–53.457
LH (miU/mL)	NOA	446	11.923	11.274–12.571
	OA	210	4.758	4.407–5.108
	Cryptozoospermia	46	6.793	5.555–8.032
	Oligoasthenozoospermia	1614	5.169	5.021–5.317
	Normal	1330	4.317	4.211–4.422
	Ejaculation disorder	6	5.5	1.688–9.312
FSH (miU/mL)	NOA	446	26.991	25.757–28.226
	OA	210	5.182	4.876–5.488
	Cryptozoospermia	46	14.693	10.891–18.496
	Oligoasthenozoospermia	1615	7.479	7.175–7.783
	Normal	1330	4.786	4.654–4.917
	Ejaculation disorder	6	10.85	-3.149–24.849
PRL (ng/mL)	NOA	446	10.474	9.987–10.961
	OA	210	9.893	9.163–10.623
	Cryptozoospermia	46	10.135	8.428–11.842
	Oligoasthenozoospermia	1614	10.271	9.762–10.780
	Normal	1330	10.997	9.262–12.732
	Ejaculation disorder	6	12.15	6.881–17.419
Testosterone (ng/mL)	NOA	447	4.021	3.716–4.325
	OA	210	4.53	4.277–4.783
	Cryptozoospermia	46	4.552	3.813–5.291
	Oligoasthenozoospermia	1616	4.804	4.711–4.896
	Normal	1330	4.937	4.843–5.031
	Ejaculation disorder	6	6.902	-2.815–16.618
E2 (pg/mL)	NOA	445	25.518	24.481–26.555
	OA	207	25.865	24.538–27.191
	Cryptozoospermia	46	28.615	23.575–33.655
	Oligoasthenozoospermia	1611	26.391	25.774–27.007
	Normal	1330	26.033	25.546–26.520
	Ejaculation disorder	6	35.217	5.199–65.235
T (ng/dL)/E2 (pg/mL)	NOA	444	16.6	15.651–17.548
	OA	207	19.288	17.913–20.663
	Cryptozoospermia	46	17.83	14.063–21.597
	Oligoasthenozoospermia	1611	20.237	19.671–20.805
	Normal	1328	20.823	20.191–21.454
	Ejaculation disorder	6	16.744	6.855–26.633
Sperm volume (mL)	NOA	445	2.877	2.705–3.048
	OA	209	2.107	1.9–2.325
	Cryptozoospermia	46	2.748	2.276–3.22
	Oligoasthenozoospermia	1618	2.808	2.725–2.89
	Normal	1333	3.017	2.933–3.1
	Ejaculation disorder	4	0.075	-0.005–0.155
Concentration ($\times 10^6$/mL)	NOA	448	0	0–0
	OA	210	0	0–0
	Cryptozoospermia	46	0.01	0.01–0.01
	Oligoasthenozoospermia	1619	29.102	26.365–31.839
	Normal	1333	79.379	44.032–114.727
	Ejaculation disorder	0	*	*

Continued

		n	Mean	95% confidence interval
Motility (%)	NOA	448	0	0–0
	OA	209	0	0–0
	Cryptozoospermia	46	15.063	8.096–22.030
	Oligoasthenozoospermia	1619	28.549	27.676–29.422
	Normal	1333	57.533	56.95–58.116
	Ejaculation disorder	1	2	*
Total sperm motility count ($\times 10^6$/mL))	NOA	448	0	0–0
	OA	210	0	0–0
	Cryptozoospermia	46	0.004	0.002–0.006
	Oligoasthenozoospermia	1619	17.014	15.55–18.479
	Normal	1333	94.328	87.598–101.059
	Ejaculation disorder	6	0	0–0

Table 3. Means of age, serum hormone levels, and semen analysis by disease state (NOA, OA, Cryptozoospermia, Oligoasthenozoospermia, Normal, and Ejaculation disorder).

Figure 2 shows the AI prediction model for the risk of male infertility generated by AutoML Tables. In its accuracy evaluation, AUC ROC and AUC PR were 74.2% and 77.2%, respectively. In a ranking of the contribution of variables from 1st to 7th, FSH came a clear 1st. The ranking order for 2nd and below was T/E2, LH, testosterone, age, E2 and PRL. The feature importance percentages for FSH, T/E2, and LH were 92.24%, 3.37%, and 1.81%, respectively. The values for Accuracy, Precision, and Recall, and the F-value were 52.2%, 49.1%, 95.8%, and 64.9%, respectively, when the threshold was 0.30. In addition, when the threshold was 0.50, the values for Accuracy, Precision and Recall, and the F-value were 71.2%, 83.0%, 47.3%, and 60.2%, respectively.

In addition, the semen analysis data from 2011 to 2020 was evaluated based on the relevant standards in the fifth edition of the WHO Manual for Human Semen Testing of 2010[13]. We defined a total motility sperm count of 9.0×10^6 ($1.5\ \text{ml} \times 15 \times 10^6/\text{ml} \times 40\%$) as the lower limit of normal, assigning a value of "0" if the total motility sperm count calculated for an individual patient was above 9.0×10^6 and a value of "1" when it was below. The AI prediction models for risk of male infertility were generated using Prediction One and AutoML Tables. In the accuracy evaluation of the AI prediction model based on Prediction One, AUC was 75.75%. In a ranking of the contribution of variables from 1st to 7th, FSH came a clear 1st. The ranking order for 2nd and below was T/E2, LH, age, testosterone, E2 and PRL. The values for Accuracy, Precision and Recall, and the F-value were 63.39%, 56.28%, 84.24%, and 67.48%, respectively, when the threshold was 0.30. In addition, when the threshold was 0.48, the values for Accuracy, Precision and Recall, and the F-value were 70.77%, 74.17%, 53.94%, and 62.46%, respectively (See Supplementary Fig. S1 online).

In the accuracy evaluation of the AutoML Tables-based model, AUC ROC and AUC PR were 82.1% and 82.3%. In a ranking of the contribution of variables from 1st to 7th, FSH came a clear 1st. The ranking order for 2nd and below was LH, T/E2, age, testosterone, E2 and PRL. The feature importance percentages for FSH, LH, and T/E2 were 73.67%, 8.63%, and 6.06%, respectively. The values for Precision, and Recall, and the F-value were 60.6%, 94.0%, and 73.7%, respectively, when the threshold was 0.30. In addition, when the threshold was 0.50, the values for Precision and Recall, and the F-value were 73.1%, 73.1%, and 73.1%, respectively (See Supplementary Fig S2 online).

We investigated the sperm analysis and serum hormone level data from 2021 and 2022 to verify the AI prediction model generated with Prediction One for the risk of male infertility. Figure 3 shows the confusion matrix in 2021 when the threshold was 0.30. The values for Accuracy, Precision, and Recall, and the F-value were 57.98%, 51.09%, 85.37%, and 63.90%, respectively. Figure 3 also shows the confusion matrix in 2022 when the threshold was 0.30. The values for Accuracy, Precision, and Recall, and the F-value were 68.07%, 70.73%, 83.65%, and 76.65%, respectively.

In 2021, thirty-five (18.62%) of the 188 cases overall were azoospermia. There were 10 cases of OA, 24 cases of NOA, and 1 case of male MHH (hypogonadotropic hypogonadism). When validated using the AI prediction model for risk of male infertility, the result for OA was 70% correct (7 cases) while the results for NOA and MHH were 100% correct (24 cases). In 2022, fifty-three (31.93%) of the 166 cases overall were azoospermia. There were 25 cases of OA and 28 cases of NOA. When validated using the AI prediction model for the risk of male infertility, the results for OA and NOA were 72% (18 cases) and 100% (28 cases) correct, respectively.

Discussion

Semen analysis is important for evaluating male infertility, and often the first test ordered when a couple presents for a fertility check or when a man is interested in permanent contraception[14–16]. Sperm motility, morphology, velocity, and concentration are investigated using microscopes and counting chambers by skilled embryologists. Other methods, such as CASA (computer-assisted semen analysis), which uses algorithms to automatically track spermatozoa, are also effective and are able to present qualitative information on sperm motility. However, semen analysis can only be done at a fertility clinic in most cases. In addition, many men feel uncomfortable about having semen analysis[11].

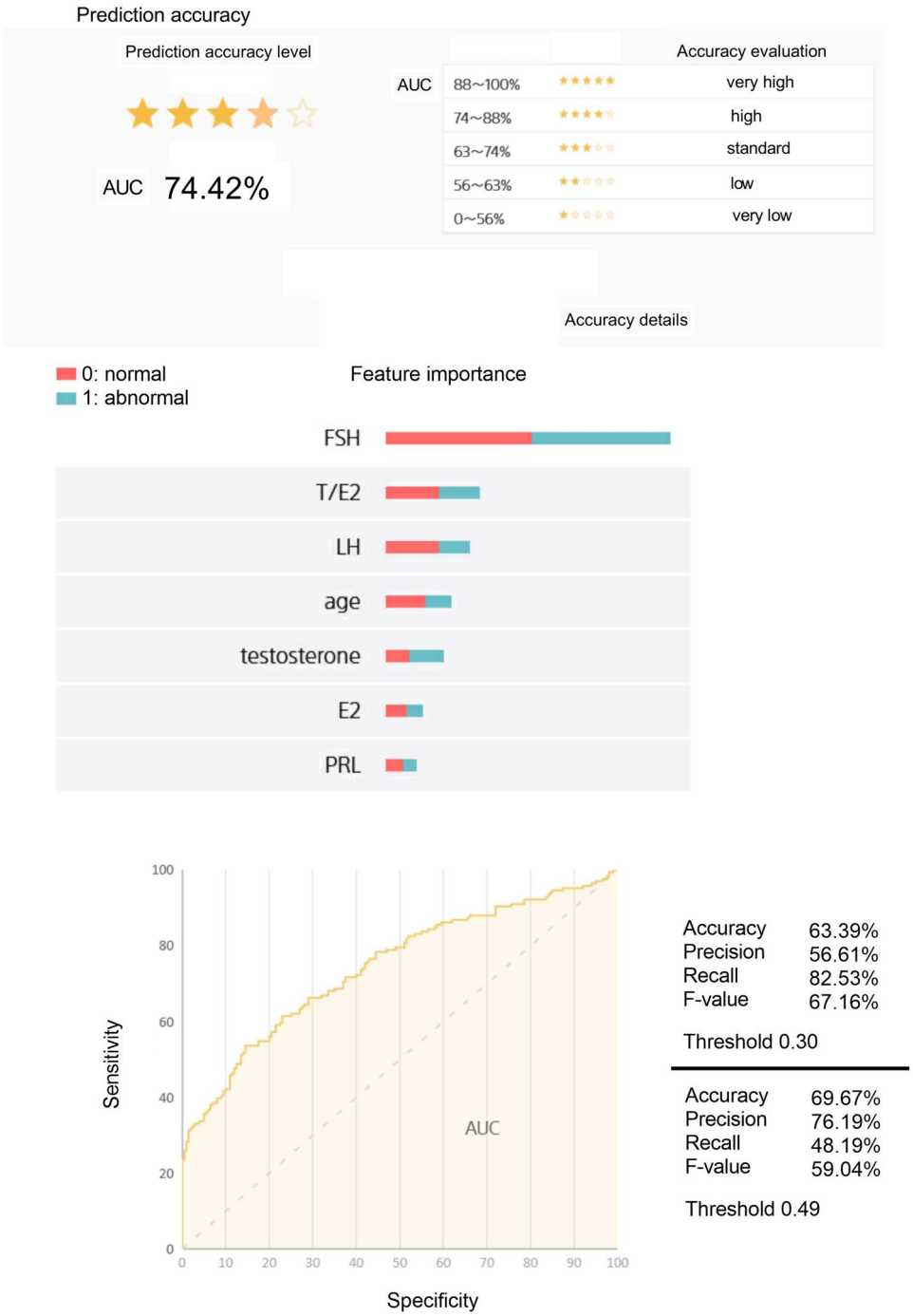


Figure 1. AI predictive analysis model using Prediction One.

Recently, the application of AI in medicine has been remarkable. Machine learning methods may improve prediction models[17,18]. We give examples of disease prediction using AI based on serum hormone levels in the following.

Although no accurate predictive models had previously been identified for hormonal prognosis in NFPA (non-functioning pituitary adenoma), Fang et al. demonstrated that machine learning models could accurately predict postoperative pituitary outcomes based on preoperative anterior pituitary hormones in NFPA[19]. In addition, it had been reported that elevated PTH (parathyroid hormone) levels are associated with higher mortality risks, and Kato et al. found that an AI model could predict elevated PTH levels among US adults. Their results suggest that even without serum calcium, phosphatase, and vitamin D levels, the model could predict elevated PTH levels[20].

We could make an AI model for evaluating male infertility using serum hormone levels, without conventional sperm analysis, and the Prediction One-based model had high accuracy as indicated by an AUC of 74.42%. An

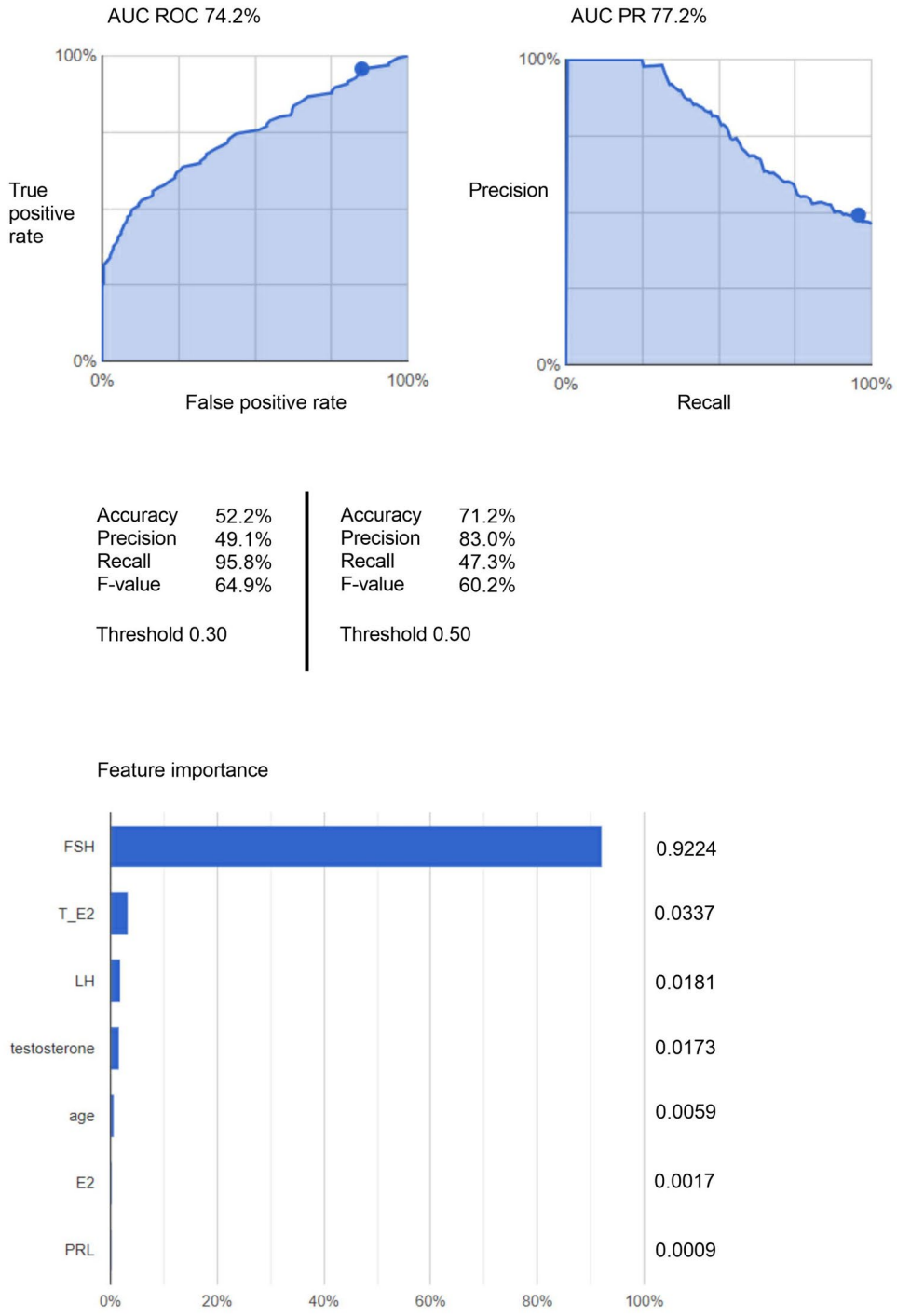


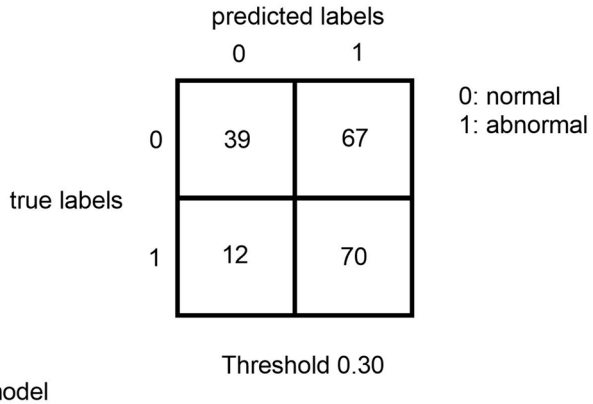
Figure 2. AI predictive analysis model using AutoML Tables.

AI prediction model with similar accuracy was created with AutoML Tables. Prediction One and AutoML Tables are both tools for automatically generating machine learning AI models, but Prediction One is designed for companies in Japan and only supports Japanese. AutoML Tables is global and supports a wide range of industries. However, it requires a minimum of 1,000 data to create an AI model. The reason for using Prediction One and AutoML Tables is that they use different machine learning algorithms and methods, so different approaches can be tried by using them. In addition, we consider that a detailed comparison of the AI models generated by Prediction One and AutoML Tables would allow us to select the AI model with the better performance. Since we also believe that Prediction One and AutoML Tables would produce AI models with comparable accuracy, this should help ensure model reliability.

The feature importance ranking of the AI models indicated that FSH had the highest importance level among all features. Its importance was significantly higher than that of "T/E2" and "LH" in the second and third

a)

Accuracy	57.98%
Precision	51.09%
Recall	85.37%
F-value	63.90%

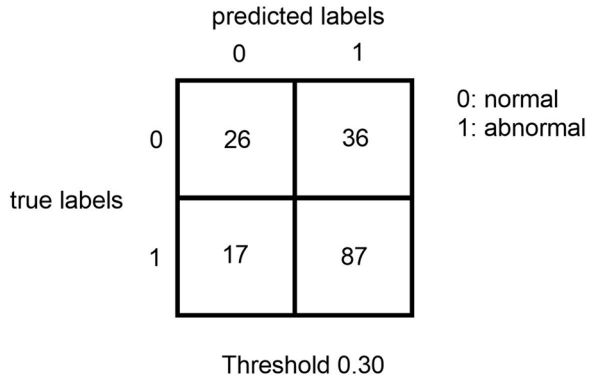


Validation of AI prediction model with azoospermia in 2021

	n	predicted labels	%
OA	10	7	70
NOA	24	24	100
MHH	1	1	100

b)

Accuracy	68.07%
Precision	70.73%
Recall	83.65%
F-value	76.65%



Validation of AI prediction model with azoospermia in 2022

	n	predicted labels	%
OA	25	18	72
NOA	28	28	100

Figure 3. Validation of AI prediction model generated by Prediction One using data from 2021 and 2022.

positions. Because the machine learning algorithms and methods used by Prediction One and AutoML Tables are different, there will be some differences in feature values. However, the feature rankings from first to third were the same for both, suggesting that the AI models generated are highly reliable.

Tradewell et al. reported a quadratic model that predicts probability of azoospermia from serum FSH levels[21]. They concluded that being able to predict the probability of azoospermia without semen analysis would be useful to urologists when counseling patients, especially when there are logistical hurdles in obtaining a formal semen analysis or for reevaluation prior to surgical sperm extraction. However, they stated that while predicting the probability of azoospermia from serum hormones will not replace semen analysis, the role of hormonal evaluation may expand with the rise of at-home diagnosis. This is also our opinion. We would like to position our AI model as a convenient means of screening for male infertility prior to semen analysis. Thus, a limitation of this study is that the AI model created is not a substitute for semen testing.

Currently, at-home diagnostics for male infertility allow men to test their semen without the bother of going to a clinic and paying a higher charge. DTC (Direct-to-consumer) home sperm kits are available from numerous companies and their use seems to be increasing at present. However, although at-home diagnostic kits for male infertility have advantages over traditional methods for semen analysis in terms of convenience and cost, they still have many limitations. First, current non-conventional sperm analysis methods are best used only for indicating whether a user should pursue further testing or not. In addition, because there is a lot of variability in

semen analysis, a single parameter does not define whether an individual is fertile or infertile. Second, at-home diagnostics for male infertility are not yet a replacement for laboratory analysis. We consider our AI model for determining risk of male infertility in patients from serum hormone levels to be superior to at-home diagnostics, since serum hormone levels are less variable than semen analysis parameters.

Meeker et al. characterized the relationship between serum hormone levels and semen quality among 388 infertile men. They defined abnormal semen concentration as $< 20 \times 10^6$/mL and found an adjusted odds ratio of 1.0 for abnormal semen concentration in men with low serum FSH levels as compared with normal serum FSH levels and a 4.6 increased odds of semen concentration $< 20 \times 10^6$/mL in men with elevated FSH levels compared to those with low FSH levels[10]. They reported that an FSH concentration greater than 10 IU/L was predictive of a sperm concentration of less than 20×10^6/mL, with a sensitivity of 0.55 (specificity = 0.79; positive predictive value = 28%)[10].

In contrast to FSH, T/E2 and LH have not received much attention regarding feature importance. When predicting the risk of male infertility using an AI model, if it is not a case of determining the likelihood of azoospermia, it would be important to evaluate not only FSH but also other features, such as T/E2 and LH. Regarding the difference between the models in predicting the two types of azoospermia, obstructive and non-obstructive, this may be related to the feature importance of T/E2 and LH, as well as that of FSH. However, we have insufficient information at present to determine a means by which AI could distinguish oligozoospermia, asthenozoospermia, teratozoospermia, and normozoospermia from obstructive azoospermia without semen analysis. Aromatase catalyzes the conversion of testosterone to E2. Aromatase inhibitors have been offered historically to patients with a T/E2 ratio < 10[22–24] and have been shown to decrease serum E2 levels and improve semen parameters in men with a low T/E2 ratio[24–26]. Therefore, it is no coincidence that T/E2 was the second largest contributor to the AI model, since T/E2 is clearly related to sperm concentration and motility.

A potential limitation of this study was the collection of a single semen sample to assess semen parameters, and the collection of a single blood sample to measure serum hormone levels, from each patient. Also, the accuracy of the AI model may be compromised because it only considers hormone levels and semen analysis but does not take history into account; for example, of varicoceles. Plymate et al. examined three groups of men: normal fertile men, fertile men with a varicocele, and infertile men with a varicocele. They found that in normal men, there was a positive correlation between serum inhibition measurements and sperm concentration and testicular volume, whereas neither group of men with varicoceles exhibited these relationships[9,27]. However, since we consider our AI model to be a preliminary screening tool in semen analysis, high accuracy would not be required. Also, since our AI model is intended for screening purposes, it prioritizes recall over precision, a measure of accuracy, with an emphasis on comprehensiveness.

Currently, male infertility is widely considered a harbinger for a man's general health and a growing body of literature has identified male infertility as a potential biomarker for both present and future health[28–30]. Salonia et al. reported that males with infertility had more medical comorbidities than fertile men[31]. Additionally, semen quality decreases as a man's medical comorbidities increase[32,33]. As a result, a man who is seeking reproductive treatment may also benefit from an evaluation of his overall health because improvements in health can also manifest as an increase in semen quality[34]. Therefore, managing male infertility has wide health ramifications. Screening based on serum hormone levels using an AI model would not only be important for evaluating male patients for infertility but also for optimizing their future health.

In conclusion, ability to predict the probability of male infertility without semen analysis would be useful to all physicians and male patients. We believe that screening for male infertility by healthcare professionals other than reproductive specialists will benefit potential male infertility patients. In future clinical application, if the AI prediction model is introduced at health check-up centers and clinical laboratory companies, for example, it will be possible to determine the risk of male infertility by only measuring serum hormone levels in adult males, without semen analysis. If abnormalities are found, individuals would be referred to an infertility facility. The model has the potential to be an unprecedented, revolutionary new tool for comprehensively identifying male infertility patients who had remained undiscovered in the past.

Methods

Study population

We retrospectively obtained data from the medical records of 3,662 patients who underwent sperm analysis and serum hormone level measurement for male infertility from 1 January 2011 to 31 December 2020. We also obtained individual data for 188 and 166 patients, respectively, on whom sperm analysis and serum hormone level measurement were performed for male infertility in 2021 and 2022. For both, the age of subjects was defined as 18 years and over. The WHO Laboratory Manual for the Examination and Processing of Human Semen of 2021[4] was used for sperm analysis. We measured serum hormonal levels of FSH, LH, PRL, total testosterone, and E2. The method of testing hormone levels was the ECLIA (electrochemiluminescence immunoassay assay). In the ECLIA, a magnetic particle-bound antibody (antigen) reacts with an antibody (antigen) labeled with a chemiluminescent substance, and then the emission from the resulting immuno-compound in a redox reaction on an electrode in the presence of TPA (Tripropylamine) is measured. In addition, the T/E2ratio was calculated using T in ng/dL and E2 in pg/mL. One of the exclusion criteria was "Underwent only semen analysis or serum hormone level measurement, not both".

The Ethics Committee of Toho University Omori Medical Center has waived informed consent for this study. The study protocol was approved by the Ethics Committee of Toho University Omori Medical Center (approval No. M22267 20,104). All methods were performed in accordance with the relevant guidelines and regulations as well as with the Declaration of Helsinki. The presented study design was accepted by the Ethics Committee on

the condition that a document declaring an opt-out policy, by which any potential patients and/or their relatives could refuse inclusion in this study, was uploaded to the website of the Toho University Omori Medical Center.

Database

Age, LH, FSH, PRL, total testosterone, E2, T/E2, sperm volume, sperm concentration, sperm motility, and total sperm motility count were extracted from patient records and Excel (Microsoft Corporation, Redmond, Washington, U.S.A) sheets were created from the data. Total sperm motility count was calculated by multiplying sperm volume by sperm concentration by sperm motility. In addition, we defined a total motility sperm count of 9.408×10^6 ($1.4\ ml \times 16 \times 10^6/ml \times 42\%$) as the lower limit of normal in accordance with WHO Laboratory Manual for the Examination and Processing of Human Semen of 2021, assigning a value of "0" if the total motility sperm count calculated for an individual patient was above 9.408×10^6 and a value of "1" when it was below.

Statistical analysis

Age, LH, FSH, PRL, total testosterone, E2, T/E2, sperm volume, sperm concentration, sperm motility, and total sperm motility count were extracted from patient records and IBM SPSS statistics software (Version 27) (IBM, Armonk, New York, U.S.A) sheets were created from the data. Statistical analysis was performed using IBM SPSS statistics software (Version 27). The mean and 95% confidence interval of each item were indicated by descriptive statistics.

Creation of machine learning prediction model requiring no coding using prediction one

Prediction One software (https://predictionone.sony.biz; Sony Network Communications Inc., Tokyo, Japan) was used to make a prediction model for determining risk of male infertility. Prediction One is only available in Japanese. It generates feature vectors from datasets using standard preprocessing methods, such as one-hot encoding for categorical variables and normalization for numerical variables. A gradient-boosting tree and a neural network are used as supervised machine learning models, each trained with hyperparameter tuning. An ensemble model of both trained models was constructed. Missing values are automatically handled by common machine learning techniques, one of them the gradient-boosting tree. The AUC was calculated using internal validation to evaluate the accuracy of the AI model. Prediction One makes the best predictive model using an artificial neural network with fivefold cross validation. It also evaluates the 'importance of variables' using a method based on permutation feature importance. This method was used to calculate the difference in the model output when a single variable was removed. The value of the difference in the model output indicated how much the model depended on the variables. The value of the difference was computed for each covariate and then averaged over those in the dataset[35].

Prediction One read in the data of the 3662 patients who underwent sperm analysis and serum hormone level measurement for male infertility and automatically divided them into internal training and cross-validation datasets, in more or less equal halves. It automatically adjusted and optimized the variables to make it easy to process them statistically and mathematically and select an appropriate algorithm with ensemble learning. The missing values were automatically compared and Prediction One made the best prediction model using an ANN (artificial neural network) with internal cross-validation. The details are trade secrets and cannot be provided.

The data from 188 and 166 patients, who underwent sperm analysis and measurement of serum hormone levels for male infertility determination in 2021 and 2022, respectively, were used as an external validation dataset. Among them, the data of those with azoospermia in 2021and 2022 were extracted and used to validate the AI prediction model for male infertility risk.

Creation of machine learning prediction model requiring no coding using AutoML tables on Google cloud platform

AutoML makes the power of machine learning available to those with limited knowledge of machine learning. AutoML is built on Google's machine learning capabilities for creating custom machine learning models. AutoML Tables enables automatic building and deployment of state-of-the-art machine learning models on structured data at massively increased speed and scale.

AutoML Tables read in the data of the 3662 patients who underwent sperm analysis and serum hormone level measurement for male infertility determination and automatically divided them into internal training and cross-validation datasets, in more or less equal halves. It automatically adjusted and optimized the variables to make it easy to process them statistically and mathematically and select an appropriate algorithm with ensemble learning. The missing values were automatically compared and AutoML Tables made the best prediction model using an ANN with internal cross-validation. The details are trade secrets and cannot be provided.

In addition, AutoML Tables indicates how much each feature impacts the model. This is shown in a feature importance graph. Values are provided as a percentage for each feature: the higher the percentage, the more strongly that feature impacts model training.

Data availability

K H has ownership for the data used with Prediction One and AutoML Tables. The data collected during this study is patient data obtained with the Ethics Committee's approval and can be shared for other research. All data generated or analyzed during this study are included in this published article.

Received: 29 February 2024; Accepted: 17 July 2024
Published online: 31 July 2024

References

1. Boivin, J., Bunting, L., Collins, J. A. & Nygren, K. G. International estimates of infertility prevalence and treatment-seeking: Potential need and demand for infertility medical care. *Hum. Reprod.* **22**, 1506–1512. https://doi.org/10.1093/humrep/dem046 (2007).
2. Fainberg, J. & Kashanian, J. A. Recent advances in understanding and managing male infertility. *F1000Res* **8**, 670 (2019).
3. Baskaran, S., Finelli, R., Agarwal, A. & Henkel, R. Diagnostic value of routine semen analysis in clinical andrology. *Andrologia* **53**, e13614. https://doi.org/10.1111/and.13614 (2021).
4. Organization, W. H. WHO laboratory manual for the examination and processing of human semen 6th edition. (2021).
5. Kathrins, M. & Niederberger, C. Diagnosis and treatment of infertility-related male hormonal dysfunction. *Nat. Rev. Urol.* **13**, 309–323. https://doi.org/10.1038/nrurol.2016.62 (2016).
6. Clavijo, R. I. & Hsiao, W. Update on male reproductive endocrinology. *Transl. Androl. Urol.* **7**, S367–S372. https://doi.org/10.21037/tau.2018.03.25 (2018).
7. Martin-du-Pan, R. C. & Bischof, P. Increased follicle stimulating hormone in infertile men. Is increased plasma FSH always due to damaged germinal epithelium?. *Hum. Reprod.* **10**, 1940–1945. https://doi.org/10.1093/oxfordjournals.humrep.a136211 (1995).
8. Sharma, A., Minhas, S., Dhillo, W. S. & Jayasena, C. N. Male infertility due to testicular disorders. *J. Clin. Endocrinol. Metab.* **106**, e442–e459. https://doi.org/10.1210/clinem/dgaa781 (2021).
9. Uhler, M. L., Zinaman, M. J., Brown, C. C. & Clegg, E. D. Relationship between sperm characteristics and hormonal parameters in normal couples. *Fertil. Steril.* **79**(Suppl 3), 1535–1542. https://doi.org/10.1016/s0015-0282(03)00336-4 (2003).
10. Meeker, J. D., Godfrey-Bailey, L. & Hauser, R. Relationships between serum hormone levels and semen quality among men from an infertility clinic. *J. Androl.* **28**, 397–406. https://doi.org/10.2164/jandrol.106.001545 (2007).
11. Nosrati, R. *et al.* Paper-based quantification of male fertility potential. *Clin. Chem.* **62**, 458–465. https://doi.org/10.1373/clinchem.2015.250282 (2016).
12. Yu, S. *et al.* Emerging technologies for home-based semen analysis. *Andrology* **6**, 10–19. https://doi.org/10.1111/andr.12441 (2018).
13. Cooper, T. G. *et al.* World Health Organization reference values for human semen characteristics. *Hum. Reprod. Update* **16**, 231–245. https://doi.org/10.1093/humupd/dmp048 (2010).
14. Guzick, D. S. *et al.* Sperm morphology, motility, and concentration in fertile and infertile men. *N. Engl. J. Med.* **345**, 1388–1393. https://doi.org/10.1056/NEJMoa003005 (2001).
15. Jouannet, P. *et al.* Study of a group of 484 fertile men. Part I: Distribution of semen characteristics. *Int. J. Androl.* **4**, 440–449. https://doi.org/10.1111/j.1365-2605.1981.tb00728.x (1981).
16. Macleod, J. & Gold, R. Z. The male factor in fertility and infertility. III. An analysis of motile activity in the spermatozoa of 1000 fertile men and 1000 men in infertile marriage. *Fertil. Steril.* **2**, 187–204 (1951).
17. Deo, R. C. Machine learning in medicine. *Circulation* **132**, 1920–1930. https://doi.org/10.1161/CIRCULATIONAHA.115.001593 (2015).
18. Rajkomar, A., Dean, J. & Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **380**, 1347–1358. https://doi.org/10.1056/NEJMra1814259 (2019).
19. Fang, Y. *et al.* Machine-learning prediction of postoperative pituitary hormonal outcomes in nonfunctioning pituitary adenomas: A multicenter study. *Front. Endocrinol. (Lausanne)* **12**, 748725 (2021).
20. Kato, H. *et al.* Machine learning-based prediction of elevated PTH levels among the US general population. *J. Clin. Endocrinol. Metab.* **107**, 3222–3230. https://doi.org/10.1210/clinem/dgac544 (2022).
21. Tradewell, M. B. *et al.* Algorithms for predicting the probability of azoospermia from follicle stimulating hormone: Design and multi-institutional external validation. *World J. Mens. Health* **40**, 600–607. https://doi.org/10.5534/wjmh.210138 (2022).
22. Gregoriou, O. *et al.* Changes in hormonal profile and seminal parameters with use of aromatase inhibitors in management of infertile men with low testosterone to estradiol ratios. *Fertil. Steril.* **98**, 48–51. https://doi.org/10.1016/j.fertnstert.2012.04.005 (2012).
23. Helo, S. *et al.* A randomized prospective double-blind comparison trial of clomiphene citrate and anastrozole in raising testosterone in hypogonadal infertile men. *J. Sex Med.* **12**, 1761–1769. https://doi.org/10.1111/jsm.12944 (2015).
24. Naelitz, B. D. *et al.* Testosterone and luteinizing hormone predict semen parameter improvement in infertile men treated with anastrozole. *Fertil. Steril.* **120**, 746–754. https://doi.org/10.1016/j.fertnstert.2023.06.032 (2023).
25. Raman, J. D. & Schlegel, P. N. Aromatase inhibitors for male infertility. *J. Urol.* **167**, 624–629. https://doi.org/10.1016/S0022-5347(01)69099-2 (2002).
26. Shoshany, O., Abhyankar, N., Mufarreh, N., Daniel, G. & Niederberger, C. Outcomes of anastrozole in oligozoospermic hypoandrogenic subfertile men. *Fertil. Steril.* **107**, 589–594. https://doi.org/10.1016/j.fertnstert.2016.11.021 (2017).
27. Plymate, S. R., Paulsen, C. A. & McLachlan, R. I. Relationship of serum inhibin levels to serum follicle stimulating hormone and sperm production in normal men and men with varicoceles. *J. Clin. Endocrinol. Metab.* **74**, 859–864. https://doi.org/10.1210/jcem.74.4.1548351 (1992).
28. Kasman, A. M., Del Giudice, F. & Eisenberg, M. L. New insights to guide patient care: The bidirectional relationship between male infertility and male health. *Fertil. Steril.* **113**, 469–477. https://doi.org/10.1016/j.fertnstert.2020.01.002 (2020).
29. Del Giudice, F. *et al.* Clinical correlation among male infertility and overall male health: A systematic review of the literature. *Investig. Clin. Urol.* **61**, 355–371. https://doi.org/10.4111/icu.2020.61.4.355 (2020).
30. Belladelli, F., Muncey, W. & Eisenberg, M. L. Reproduction as a window for health in men. *Fertil. Steril.* **120**, 429–437. https://doi.org/10.1016/j.fertnstert.2023.01.014 (2023).
31. Salonia, A. *et al.* Are infertile men less healthy than fertile men? Results of a prospective case-control survey. *Eur. Urol.* **56**, 1025–1031. https://doi.org/10.1016/j.eururo.2009.03.001 (2009).
32. Eisenberg, M. L., Li, S., Behr, B., Pera, R. R. & Cullen, M. R. Relationship between semen production and medical comorbidity. *Fertil. Steril.* **103**, 66–71. https://doi.org/10.1016/j.fertnstert.2014.10.017 (2015).
33. Ventimiglia, E. *et al.* Infertility as a proxy of general male health: Results of a cross-sectional survey. *Fertil. Steril.* **104**, 48–55. https://doi.org/10.1016/j.fertnstert.2015.04.020 (2015).
34. Shiraishi, K. & Matsuyama, H. Effects of medical comorbidity on male infertility and comorbidity treatment on spermatogenesis. *Fertil. Steril.* **110**(1006), 1011e1002. https://doi.org/10.1016/j.fertnstert.2018.07.002 (2018).
35. Mazaki, J. *et al.* A novel predictive model for anastomotic leakage in colorectal cancer using auto-artificial intelligence. *Anticancer Res.* **41**, 5821–5825. https://doi.org/10.21873/anticanres.15400 (2021).

Acknowledgements

K H has received support in the form of a Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science (JSPS) (JSPS KAKENHI Grant Number JP22K09486). We thank CreaTact Inc. for evaluation and checking of the data of 3,662 patients who underwent sperm analysis and serum hormone level measurement. We are particularly grateful to Alexander Cox for his painstaking work as medical editor.

Author contributions

All authors contributed to the conception and design of the study. K.H. collected the data to build an AI model from clinical records. K.H. provided training in the automated machine learning models. K.H. drafted the manuscript.

Funding

This work is supported by a Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science (JSPS) (JSPS KAKENHI Grant Number JP22K09486).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-67910-0.

Correspondence and requests for materials should be addressed to H.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.