A cloud-based flexible solution for psychometric tests validation, administration and evaluation

Grazia Lo Sciuto^a, Samuele Russo^b and Christian Napoli^c

^{a} Department of Electrical, Electronics and Computer Engineering, University of Catania
Viale Andrea Doria 6, Catania 95126 CT, Italy
glosciuto@dii.unict.it

^{b} Advanced specialization student, University of Palermo
Piazza Marina 61, Palermo 90133 PA, Italy
samuelerussoct@gmail.com

^{c} Department of Mathematics and Computer Science, University of Catania
Viale Andrea Doria 6, Catania 95126 CT, Italy
napoli@dmi.unict.it

Abstract—**Psychological tests generally provide an evaluation scale to evaluate whether or not the subject manifests some traits. Such kind of tests are generally used for attitude evaluation, personal selection, educational and rehabilitation purposes, as well as for the diagnosis of cognitive disorders. The use of test and other questions-based diagnostic tools, represents one of the main and principal actions in order to start a clinical and therapeutic path, as well as for the evaluation and assessment of the possible educational and rehabilitation effort. Unfortunately such tests are generally the result of a long and difficult process of validation for their standardization, simplification and reorganizations driven by operations performed by means of complex statistical methods. In the work presented on this paper we developed a unified cloud-based resource for the management and execution of all the task related to psychometric testing, from the creation of a test, to its validation and use. The solution has been designed to grant maximum flexibility allocating resources on a cloud service. Such resources can be used as a remote support for the psychologists designing and administering the test, as well as computing platform to unburden the single terminals of the heavy computations required during the standardization procedure. Moreover, by means of the distributed database, our solution is also able to support the simplification and reorganization process, as well as to serve as online platform for the administration and consequent scoring of the finalized and standardized test.**

Index Terms—**component, formatting, style, styling, insert**

I. Introduction

In the recent years a ever increasing number of responsibilities has been assigned to expert on the field of mental health and psychology. Moreover the number of actual applications for such a professional group is continuously growing. Besides the classical occupations in the clinical environment, for both public and private practice, psychologists are nowadays employed in many different contests such as human resource selection, hiring and management, as well as on the field of education and rehabilitation. In all the said application a psychologist cannot avoid to use highly specific diagnostic tools such as psychological and psychometric tests [1].

The use of test and other questions-based diagnostic tools, represents one of the main and principal actions in order to start a clinical and therapeutic path, as well as for the evaluation and assessment of the possible educational and rehabilitation effort [2].

Moreover these kind of tools are nowadays largely used also for human resource management, as well as for personal selection and assessment before hiring. In clinical psychology the assessment phase can be based on the submission of several tests batteries in order to formulate a diagnostic hypothesis based on the answers given by the psychological patient [3]. The administered questionnaires can vary both in length and structure, on the other hand a great majority of such tests are given in multiple-answers form.

Psychological tests generally provide an evaluation scale to evaluate whether or not the subject manifests some traits. A classical example can be the Mini-Mental Test [**?**], la Symptom Checklist-90 [4], il Minnesota multiphasic personality inventory [5], [6], etc. Among the most widely used assessment tools it is possible to notice several psychometric tests for cognitive skills assessment and evaluation. Such kind of tests are generally used for attitude evaluation, personal selection, educational and rehabilitation purposes, as well as for the diagnosis of cognitive disorders. The most known tests of this kind are the Wechsler Adult Intelligence Scale [7], the Wechsler Intelligence Scale for Children [8], [9], the Raven's progressive matrices and vocabulary scales [10], etc...

Unfortunately such tests are generally the result of a long and difficult process of validation for their standardization,
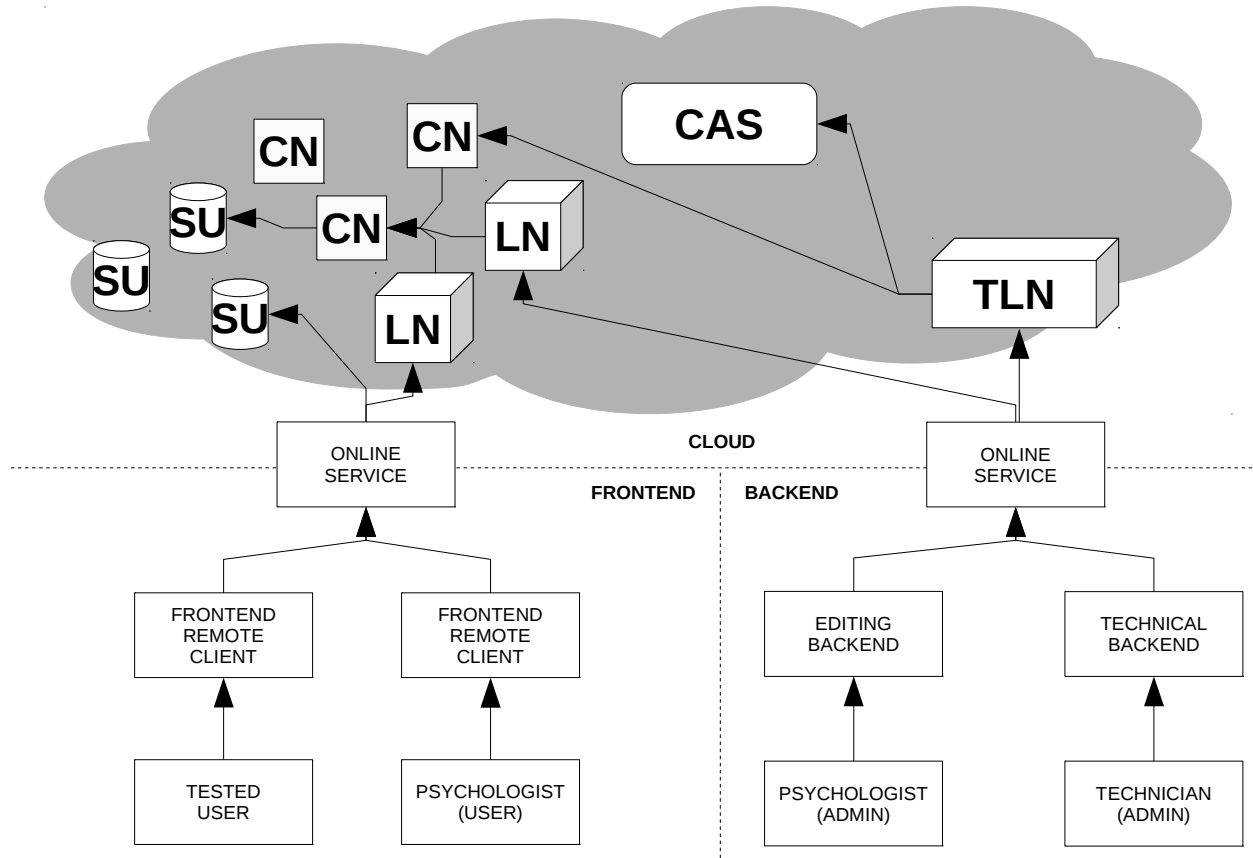
Fig. 1: A general schema of the system developed in this work. Clockwise: a representation of the cloud and its resources, the administrative and test design backend, and, finally, the users' frontend.

simplification and reorganization. Generally such operations are performed by means of complex statistical methods which make use of heavy computations (i.e. very large covariance matrices [11], factorial analysis [12], multidimensional scaling [13], cluster analysis [14], etc...).

In the work presented on this paper we developed a unified cloud-based resource for the management and execution of all the task related to psychometric testing, from the creation of a test, to its validation and use. The solution has been designed to grant maximum flexibility allocating resources on a cloud service. Such resources can be used as a remote support for the psychologists designing and administering the test, as well as computing platform to unburden the single terminals of the heavy computations required during the standardization procedure. Moreover, by means of the distributed database, our solution is also able to support the simplification and reorganization process, as well as to serve as online platform for the administration and consequent scoring of the finalized and standardized test.

The paper is organized as follows. After this brief introduction, in the following Section II the designed system is described in its constituent parts. In Section III we will focus

on the management of the cloud services giving further details on the resource allocation policies. Finally in Section IV we will report a pilot case study and the obtained results. Finally in Section V we will draw our conclusions.

II. THE DEVELOPED SYSTEM

In Figure 1 a gross schema of the designed system is reported, this is composed by the following agents and components:

A. Frontend:
- *Online interface*
- *Frontend remote clients*

B. Backend:
- *Online interface*
- *Administration backend*
- *Technical backend*

C. Cloud
- *Trusted Login Node* (TLN)
- *Cloud administration service* (CAS)
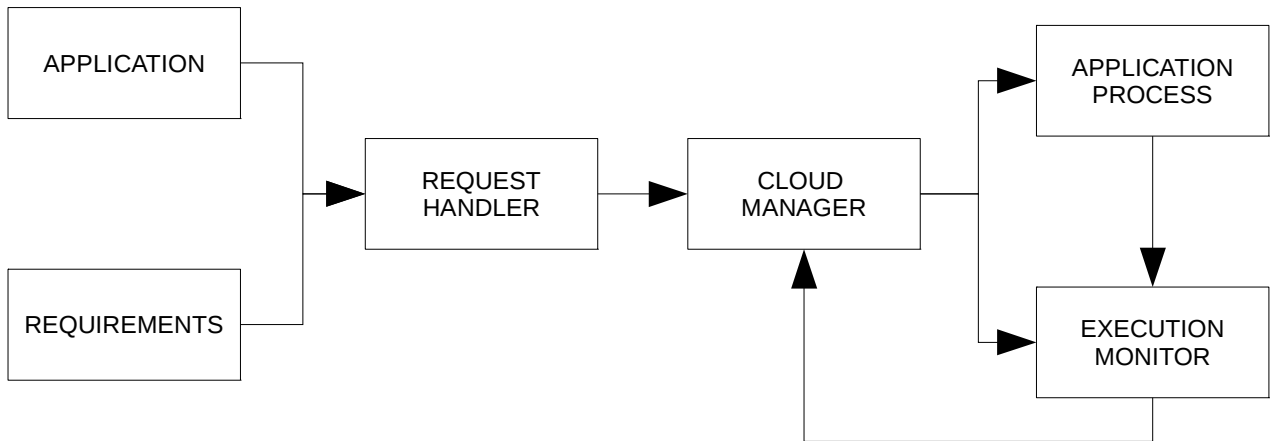- *Login nodes* (LN)
- *Computing nodes* (CN)

Fig. 2: Schematics of the Cloud Administration Service (CAS).

- *Storage Units* (SU)

The components are better described in the following.

A. Frontend

The frontend of the system has been developed by means of the Angular JS [15], [16] framework in order to grant portability and compatibility with almost all the available hardware and software systems. In this manner there are no particular requirements to interface with the developed system, granted the ability to execute JavaScript on a browser-like application. Although a web browser would have sufficed to interface with the online service, we developed a simple ad-hoc application to oversimplify the interface. In this manner it is possible to avoid unnecessary distractions during the test execution. Finally a psychologist provided with the necessary credentials can log into the system to administer the test to a patient once such a test has been standardized and approved to be used. The *frontend remote client* only provides the interface for the final *users*, but it is not designed to create a new tests or operate for its standardization, since these latter procedures are related with the backend.

B. Backend

While the frontend for the proposed system is constituted by the final interface that the users and patients can use for the only purpose of executing a certain test, the the backend of the developed system provides the necessary support for the design of new tests and their standardization as well as for the technical administration of the overall platform. Differently from the frontend, the backend provides two separated consoles for psychologists and technicians. The first allows the psychologist to design a new test, insert the item, provide the validation rules, and request to the system to analyze the validation data. The second console is reserved for technical administration in terms of resource allocation, cloud management policies, etc...

C. Cloud

The cloud resources are allocated both for computational and provisional purposes. Complete standardized tests can be administered and performed by means of the frontend interface. In this case a simple set of queries can do the job, by selecting and extracting the required data from the databases, distributed on several *storage units* (SU), as well as by uploading the given answers for further use. On the other hand the standardization of a new test, due to the required statistical analysis, also makes use of the *computing nodes* (CN). The cloud system is also provided with several login node in order to avoid direct interactions with the allocated computing nodes and storage units, excepted for the storage units containing the public database useful to run the fronted interface. Finally the technical administration of the cloud, due to the criticality of the matter, makes use of a *trusted login node* that ensure a grater security level. The details on the cloud policies are given in the following Section III.

III. THE CLOUD ENVIRONMENT

In our proposal the cloud environment is administered on a technical level by expert users. Moreover the system has been studied in order to adapt to the total load by allocating or freeing resources. For the implementation we supported our system with the Amazon Web Services (AWS) [17], and particularly on the AWS ECS and S3 service [18]. The main component for the administration of the cloud environment is the Cloud Administration Service (CAS) depicted in Figure 2. The Cloud Administration Service analyzes the provided application and estimate the computational burden by means of a XML application requirement descriptor. Along with the single application (e.g. a meta-analysis by means of factorial analysis, multidimensional scaling, etc...), the psychologist administrator also submits a set of requirements (e.g. the desired deadline or throughput, etc...). Both the application descriptor and the submitted requirements, are then analyzed
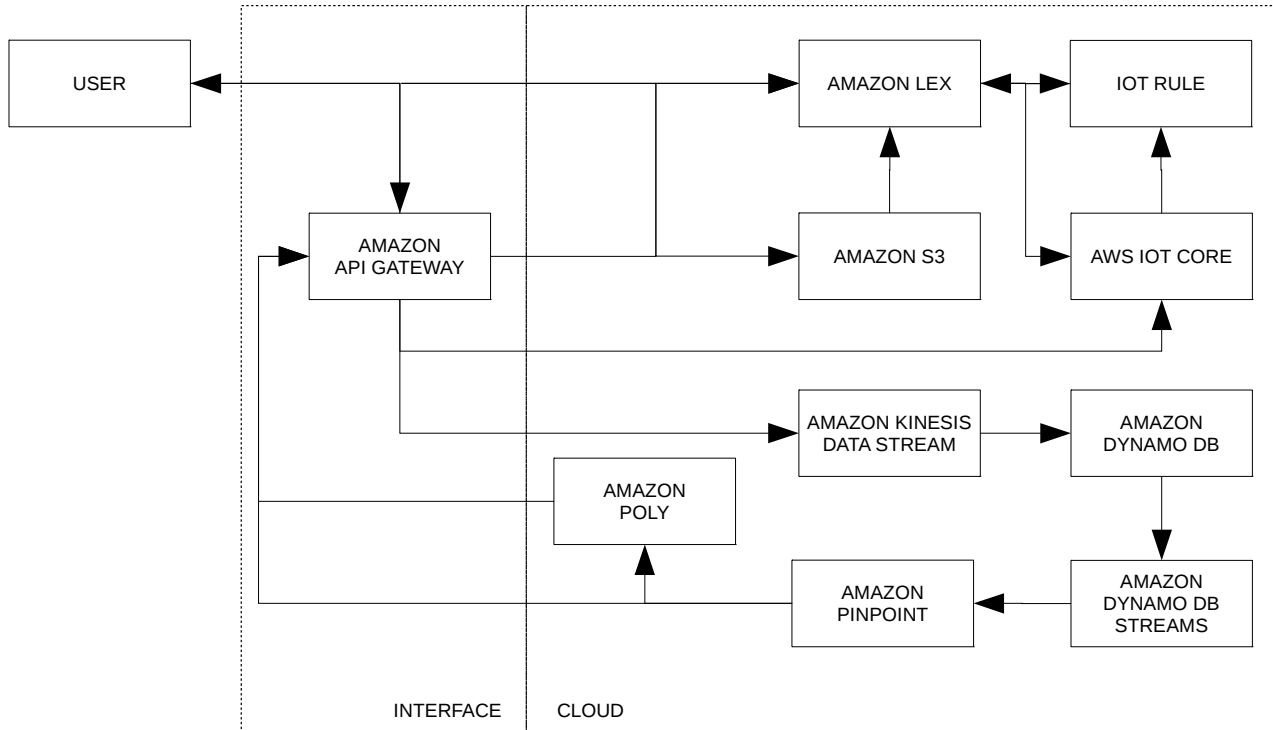
Fig. 3: The adopted Amazon Web Services (AWS) configuration and the relative data flow among the different component and services within the cloud environment

.

by the *request handler* module. The request handler has the responsibility to determine the correct allocation request mediating between the application requirements, the user defined constraints, and the effective resource availability on the cloud system (see Figure 3).

The resource request is provided to the *cloud manager* component which uses the Amazon AWS APIs to effectively request the allocation of new virtual machines. The cloud manager interfaces with the *AWS IoT Core* taking into consideration the *AWS IoT rule* component that determine the policies for the *Amazon Kinesis Data Stream*. The Amazon Kinesis Data Stream is a real-time streaming service that provides event-driven messaging and supports extended microservice architectures. This latter allows the processing requests trough the *Amazon API Gateway* once an admin has been logged and identified trough his credentials by the *Amazon Lex* component to access the *Amazon S3* service.

In our system design also the database is distributed on the cloud and supported by the *Amazon DynamoDB* services that allows data flow by means of the *Amazon DynamoDb Streams* component. Data transactions and session state are encrypted at-rest and securely managed in the high-performance and scalable NoSQL datastore offered by DynamoDB. The Amazon DynamoDB Streams is also able to trigger an *AWS Lambda function* in order to send notifications, by means of the

Amazon Pinpoint and *Amazon Polly* services, to psychologist users when a patient has completed a test, as well as to send notification to a psychologist admin when the validation and standardization process advances or changes status.

An example of the system in action is provided in the following Section IV.

IV. A CASE STUDY

In order to check the developed system on the field, we simulated the development of a novel test for the diagnosis of Attention deficit and hyperactivity disorder (ADHD). We mixed two well known and largely used test: the Conners Rating Scales for Teacher (CRST) [19] and the SCOD [20]. Among the available items we selected the following 54 items concerning the related subscales:

CRST (36 items):
- Oppositional-defiant disorders (10 items)
- Conduct disorders (5 items)
- Psychosomatic disorders (6 items)
- Anxiety and shyness (8 items)
- Perfectionism (6 items)

SCOD (18 items):
- Attention deficit (9 items)
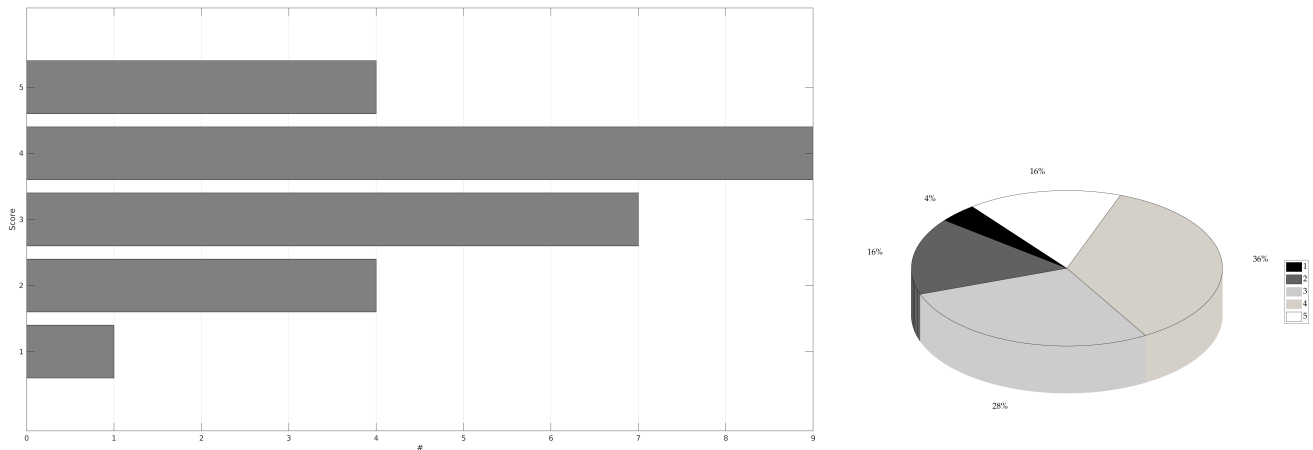- Hyperactivity (9 items)

Fig. 4: Experts' scoring of the developed system

The draft of such a test has been submitted to the system for the validation sequence, after it it has been undergone a standardization procedure that reduced it to 30 items. After that the so-created 32-items test has been used on a group of non-psychologist volunteers. The results has been made available in the system to be securely viewed by the psychologist involved in the system check.

While the reported case is only an example, the system has been tested with the help of 25 psychologists that, after using the developed system, evaluated the overall performances and utility as an asset for their profession.Figure 4 shows the results of such a poll that indicated an high degree of appreciation for the developed solution with 80% of good evaluations among the overall received scores.

V. CONCLUSION

Test and other questions-based diagnostic tools that represents one of the main and principal actions in order to start a clinical and therapeutic path, as well as for the evaluation and assessment of the possible educational and rehabilitation effort, are generally the result of a long and difficult process of validation for their standardization, simplification and reorganizations driven by operations performed by means of complex statistical methods. In this paper we have shown the utility of a novel unified cloud-based resource for the management and execution of all the task related to psychometric testing, from the creation of a test, to its validation and use. The solution has been designed to grant maximum flexibility allocating resources on a cloud service. It has been shown that such resources can be used as a remote support for the psychologists designing and administering the test, as well as computing platform to unburden the single terminals of the heavy computations required during the standardization procedure. Moreover, by means of the distributed database, our solution has been shown to be able to support the simplification and reorganization process, as well as to serve as online platform for the administration and consequent scoring of the finalized and standardized test. The system has been tested with the help of 25 psychologists that, after using the developed system, evaluated the overall performances and utility as an asset for their profession with an high degree of appreciation for the developed solution.

REFERENCES

[1] David Watson and Lee Anna Clark. Clinical diagnosis at the crossroads. *Clinical Psychology: Science and Practice*, 13(3):210–215, 2006.
[2] Max Gluckman. *Clinical psychology: The study of personality and behavior*. Routledge, 2017.
[3] Andrew M Pomerantz. *Clinical psychology: Science, practice, and culture*. Sage Publications, 2016.
[4] Leonard R Derogatis and Rachael Unger. Symptom checklist-90-revised. *The Corsini encyclopedia of psychology*, pages 1–2, 2010.
[5] Starke Rosecrans Hathaway and John Charnley McKinley. Minnesota multiphasic personality inventory; manual, revised. 1951.
[6] James N Butcher. Minnesota multiphasic personality inventory. *The Corsini Encyclopedia of Psychology*, pages 1–3, 2010.
[7] David Wechsler. Wechsler adult intelligence scale–fourth edition (wais–iv). *San Antonio, TX: NCS Pearson*, 22:498, 2008.
[8] David Wechsler. Wechsler intelligence scale for children. 1949.
[9] David Wechsler. *Wechsler preschool and primary scale of intelligence-fourth edition*. The Psychological Corporation San Antonio, TX, 2012.
[10] John Carlyle Raven et al. *Raven's progressive matrices and vocabulary scales*. Oxford pyschologists Press, 1998.
[11] Donald A Rock, Charles E Werts, and Ronald L Flaugher. The use of analysis of covariance structures for comparing the psychometric properties of multiple variables across populations. *Multivariate Behavioral Research*, 13(4):403–418, 1978.
[12] Cyril Burt. The factorial analysis of qualitative data. *British Journal of Statistical Psychology*, 3(3):166–185, 1950.
[13] Ingwer Borg and Patrick Groenen. Modern multidimensional scaling: Theory and applications. *Journal of Educational Measurement*, 40(3):277–280, 2003.
[14] Leonard Kaufman and Peter J Rousseeuw. *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons, 2009.
[15] Brad Green and Shyam Seshadri. *AngularJS*. " O'Reilly Media, Inc.", 2013.

[16] Peter Bacon Darwin and Pawel Kozlowski. *AngularJS web application development*. Packt Publ., 2013.

[17] James Murty. *Programming amazon web services: S3, EC2, SQS, FPS, and SimpleDB*. " O'Reilly Media, Inc.", 2008.

[18] Zhiwu Xie, Yinlin Chen, Tingting Jiang, Julie Speer, Tyler Walters, Pablo A Tarazaga, and Mary Kasarda. On-demand big data analysis in digital repositories: A lightweight approach. In *International Conference on Asian Digital Libraries*, pages 274–277. Springer, 2015.

[19] C Keith Conners. Conners 3. MHS, 2008.

[20] Gian Marco Marzocchi, Jaap Oosterlaan, Tiziana De Meo, Mario Di Pietro, Sara Pezzica, Pina Cavolina, JA Sergeant, and Alessandro Zuddas. Scala di valutazione dei comportamenti dirompenti per insegnanti (scod-i): validazione e standardizzazione di un questionario per la valutazione dei comportamenti dirompenti a scuola. *Giornale di Neuropsichiatria dellEtà Evolutiva*, 21:378–393, 2001.

[21] John H Wasson, Harold C Sox, Raymond K Neff, and Lee Goldman. Clinical prediction rules: applications and methodological standards. *New England Journal of Medicine*, 313(13):793–799, 1985.

[22] Emily Goodwin, Gisli H Gudjonsson, Jon Fridrik Sigurdsson, and Susan Young. The impact of adhd symptoms on intelligence test achievement and speed of performance. *Personality and Individual Differences*, 50(8):1273–1277, 2011.

[23] Gale H Roid and Lucy J Miller. Leiter international performance scale-revised (leiter-r). *Wood Dale, IL: Stoelting*, 1997.

[24] Ariane Sroubek, Mary Kelly, and Xiaobo Li. Inattentiveness in attention-deficit/hyperactivity disorder. *Neuroscience bulletin*, 29(1):103–110, 2013.

[25] Robert Wood and Albert Bandura. Impact of conceptions of ability on self-regulatory mechanisms and complex decision making. *Journal of personality and social psychology*, 56(3):407, 1989.

[26] Geert-Jan Geersing, Kristel J Janssen, Ruud Oudega, Henk van Weert, Henri Stoffers, Arno Hoes, Karel Moons, AMUSE Study Group, et al. Diagnostic classification in patients with suspected deep venous thrombosis: physicians' judgement or a decision rule? *Br J Gen Pract*, 60(579):742–748, 2010.

[27] Luiz Moutinho, Paulo Rita, and Shuliang Li. Strategic diagnostics and management decision making: a hybrid knowledge-based approach. *Intelligent Systems in Accounting, Finance & Management: International Journal*, 14(3):129–155, 2006.

[28] John A Swets, Robyn M Dawes, and John Monahan. Psychological science can improve diagnostic decisions. *Psychological science in the public interest*, 1(1):1–26, 2000.

[29] Pat Croskerry. A universal model of diagnostic reasoning. *Academic medicine*, 84(8):1022–1028, 2009.

[30] Alvan R Feinstein. *Clinimetrics*. Yale University Press, 1987.

[31] Ian D Graham, Ian G Stiell, Andreas Laupacis, Annette M O'Connor, and George A Wells. Emergency physicians' attitudes toward and use of clinical decision rules for radiography. *Academic Emergency Medicine*, 5(2):134–140, 1998.

[32] Nathan M Finnerty, Robert M Rodriguez, Christopher R Carpenter, Benjamin C Sun, Nik Theyyunni, Robert Ohle, Kenneth W Dodd, Elizabeth M Schoenfeld, Kendra D Elm, Jeffrey A Kline, et al. Clinical decision rules for diagnostic imaging in the emergency department: a research agenda. *Academic Emergency Medicine*, 22(12):1406–1416, 2015.

[33] David E Newman-Toker and Jonathan A Edlow. High-stakes diagnostic decision rules for serious disorders: the ottawa subarachnoid hemorrhage rule. *JAMA*, 310(12):1237–1239, 2013.

[34] NA Korenevskiy. Application of fuzzy logic for decision-making in medical expert systems. *Biomedical Engineering*, 49(1):46–49, 2015.

[35] Ralf HJM Kurvers, Stefan M Herzog, Ralph Hertwig, Jens Krause, Patricia A Carney, Andy Bogart, Giuseppe Argenziano, Iris Zalaudek, and Max Wolf. Boosting medical diagnostics by pooling independent judgments. *Proceedings of the National Academy of Sciences*, 113(31):8777–8782, 2016.

[36] Terrence M Shaneyfelt, Michael F Mayo-Smith, and Johann Rothwangl. Are guidelines following guidelines?: The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *Jama*, 281(20):1900–1905, 1999.

[37] Kathleen N Lohr, Marilyn J Field, et al. *Guidelines for clinical practice: from development to use*. National Academies Press, 1992.

[38] Cynthia Madden, Donald B Witzke, Arthur B Sanders, John Valente, and Mark Fritz. High-yield selection criteria for cranial computed tomography after acute trauma. *Academic Emergency Medicine*, 2(4):248–253, 1995.

[39] Michael D Cabana, Cynthia S Rand, Neil R Powe, Albert W Wu, Modena H Wilson, Paul-Andre C Abboud, and Haya R Rubin. Why don't physicians follow clinical practice guidelines?: A framework for improvement. *Jama*, 282(15):1458–1465, 1999.

[40] Erna Kentala, Ilmari Pyykkö, Kati Viikki, and Martti Juhola. Production of diagnostic rules from a neurotologic database with decision trees. *Annals of Otology, Rhinology & Laryngology*, 109(2):170–176, 2000.

[41] Wei-Yin Loh. Classification and regression trees. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):14–23, 2011.

[42] Tin Kam Ho. Random decision forests. In *Proceedings of 3rd international conference on document analysis and recognition*, volume 1, pages 278–282. IEEE, 1995.

[43] Simon Bernard, Laurent Heutte, and Sebastien Adam. On the selection of decision trees in random forests. In *2009 International Joint Conference on Neural Networks*, pages 302–307. IEEE, 2009.

[44] Jehad Ali, Rehanullah Khan, Nasir Ahmad, and Imran Maqsood. Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI)*, 9(5):272, 2012.

[45] Brendan M Reilly and Arthur T Evans. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Annals of internal medicine*, 144(3):201–209, 2006.

[46] Andreas Laupacis, Nandita Sekar, et al. Clinical prediction rules: a review and suggested modifications of methodological standards. *Jama*, 277(6):488–494, 1997.

[47] Michael J Fine, Thomas E Auble, Donald M Yealy, Barbara H Hanusa, Lisa A Weissfeld, Daniel E Singer, Christopher M Coley, Thomas J Marrie, and Wishwa N Kapoor. A prediction rule to identify low-risk patients with community-acquired pneumonia. *New England journal of medicine*, 336(4):243–250, 1997.

[48] Ian G Stiell, Gary H Greenberg, R Douglas McKnight, Rama C Nair, I McDowell, and James R Worthington. A study to develop clinical decision rules for the use of radiography in acute ankle injuries. *Annals of emergency medicine*, 21(4):384–390, 1992.

[49] Clement J McDonald and J Marc Overhage. Guidelines you can follow and can trust: an ideal and an example. *Jama*, 271(11):872–873, 1994.

[50] John H Wasson and Harold C Sox. Clinical prediction rules: have they come of age? *Jama*, 275(8):641–642, 1996.

[51] American Psychiatric Association et al. *Diagnostic and statistical manual of mental disorders, fifth edition (DSM-5®)*. American Psychiatric Pub, 2013.

[52] Tetsuo Ashizawa, Cynthia Gagnon, William J Groh, Laurie Gutmann, Nicholas E Johnson, Giovanni Meola, Richard Moxley, Shree Pandya, Mark T Rogers, Ericka Simpson, et al. Consensus-based care recommendations for adults with myotonic dystrophy type 1. *Neurology: Clinical Practice*, 8(6):507–520, 2018.

[53] Marshal F Folstein, Susan E Folstein, and Paul R McHugh. mini-mental state: a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 12(3):189–198, 1975.

[54] William A Hunt. The future of diagnostic testing in clinical psychology. *Journal of Clinical Psychology*, 1946.