

CADERNOS DE METROLOGIA

**Especial Medições para a Saúde e
Inteligência Artificial/2021**



2021 Instituto Nacional de Metrologia, Qualidade e Tecnologia (Inmetro)

Os autores são responsáveis pela apresentação dos fatos contidos e opiniões expressas nesta obra.

Ministro da Economia: Paulo Guedes

Secretário Especial de Produtividade, Emprego e Competitividade: Carlos Alexandre da Costa

Presidente do Inmetro: Marcos Heleno Guerson de Oliveira Junior

Diretor de Metrologia Científica e Tecnologia: Beniamin Achilles Bondarczuk

Editores: Aline O. Coelho, Eveline De Robertis, e Rafael de Oliveira Lattanzi Vaz

Diagramação: Eveline De Robertis

cadernodemetrologia@inmetro.gov.br

Ficha catalográfica elaborada pela Biblioteca do Inmetro

C122 Cadernos de Metrologia [recurso eletrônico] / Diretoria de Metrologia Científica e Tecnologia do Instituto Nacional de Metrologia, Qualidade e Tecnologia. N. 2 (2021). – Brasília, DF : Inmetro, 2021.

Irregular.

Modo de acesso: <https://www.gov.br/inmetro/pt-br/centrais-de-conteudo/publicacoes/documentos-tecnicos-em-metrologia> (este exemplar).

1. Metrologia - Periódicos. I. Instituto Nacional de Metrologia, Qualidade e Tecnologia.

CDD. 389.1

Inmetro
Centro Empresarial Parque Brasília
SIG, Quadra 1, Lote 985, 1º andar, Setor de Indústrias
Gráficas, Brasília – DF
CEP 70610-410
Tel +55 (61) 3348-6300
www.gov.br/inmetro/pt-br

Sumário

1

Editorial



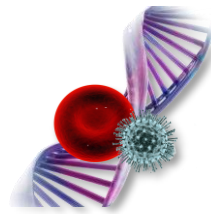
2



Saúde 4.0 e os desafios para a Metrologia Legal - Rafael Feldmann Farias & Ana Gleice Santos

10

Desafios da contagem de células humanas na biometrologia - Daniella Paiva Bonfim & Fernanda Leve



16



Princípios das Boas Práticas de Laboratório (BPL) e o Grau de Percepção por Estudantes e Pesquisadores da Biotecnologia – Elisa Rosa dos Santos e José Mauro Granjeiro

27

Inteligência Artificial e Metrologia – Charles B Prado



30



Big Data, Artificial Intelligence e Machine Learning no contexto da Análise de Dados em Data Science – Hércules de Melo Barcelos & Werickson Rocha

40

Inteligência Artificial Aplicada a Metrologia - Uma ferramenta de Calibração - Elizeu Calegari, Werickson Rocha & Gabriel Sarmanho



Editorial

Esta edição dos Cadernos de Metrologia é duplamente especial: além de tratar do tema do **Dia Mundial da Metrologia** deste ano, **Medições para a saúde**, dedica uma seção sobre **Inteligência Artificial** (tema da Semana Nacional de Ciência e Tecnologia de 2020), uma das principais pautas desses novos tempos.

O momento em que vivemos, e a pandemia de Covid-19 ainda a assolar o planeta, tornam ainda mais necessário destacar o importante papel que a medição desempenha na saúde e no bem-estar de cada um de nós. Ela está na descoberta científica e na inovação, na fabricação industrial e no comércio internacional, na melhoria da qualidade de vida e na proteção do meio ambiente. A Metrologia está presente na produção de testes mais eficazes, de equipamentos de medição necessários para o desenvolvimento de vacinas, na colaboração com outras instituições para a caracterização de patógenos, entre outros.

Nesta primeira parte dos Cadernos de Metrologia, encontramos os artigos Saúde 4.0 e os desafios para a Metrologia Legal, de Rafael Feldmann Farias e Ana Gleice Santos; Desafios da contagem de células humanas na biometrologia, de Daniella Paiva Bonfim e Fernanda Leve; e Princípios das Boas Práticas de Laboratório (BPL) e o grau de percepção por estudantes e pesquisadores da biotecnologia, por Elisa Rosa dos Santos e José Mauro Granjeiro.

Seção especial – Inteligência artificial

A **Inteligência Artificial** é um tema que representa tanto a potencialidade científica de produzir máquinas autônomas cada vez mais inteligentes quanto a busca humana pela compreensão de sua própria natureza. A área traz grandes desafios e oportunidades de solucionar problemas de natureza econômica, tecnológica e social, entre outros. O Inmetro está ciente e atuante nessa agenda, o que podemos conferir nos artigos Inteligência Artificial e Metrologia, por Charles Prado; Big Data, Artificial Intelligence e Machine Learning no contexto da Análise de Dados em Data Science, de Hércules de Melo Barcelos e Werickson Rocha; e Inteligência Artificial aplicada a Metrologia - Uma ferramenta de Calibração, de Elizeu Calegari, Werickson Rocha e Gabriel Sarmanho.

Esta edição foi preparada com o propósito de disseminar a cultura da Metrologia e o importante trabalho dos pesquisadores do Inmetro. Por isso, envie suas impressões e sugestões para cadernodemetrologia@inmetro.gov.br

Esperamos que apreciem, boa leitura!

Os Editores



Saúde 4.0 e os desafios para a Metrologia Legal



Rafael F Farias

Diretoria de Metrologia Legal (Dimel)

E-mail: rffarias@inmetro.gov.br

Ana Gleice Santos

Diretoria de Metrologia Legal (Dimel)

O que é Saúde 4.0?

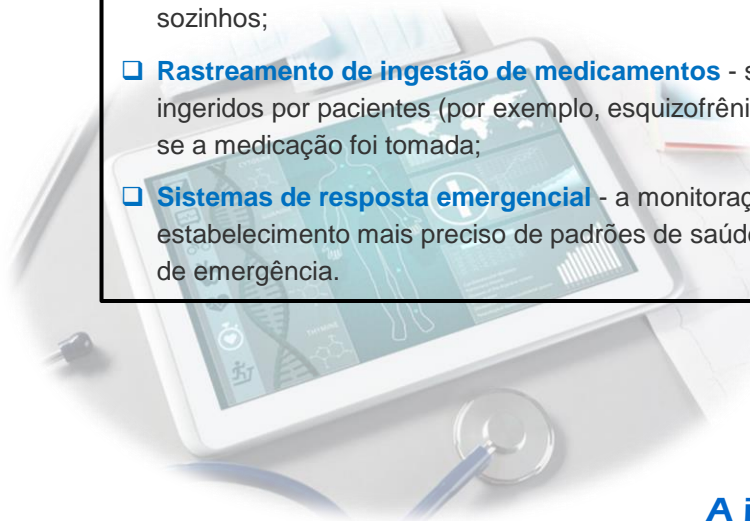
Em 2005, chegava às telas dos cinemas o filme “A ilha” que tem como tema central o debate ético em torno da clonagem. A despeito disso, uma cena bastante curiosa passa despercebida: o protagonista Lincoln 6E utiliza o banheiro e, instantes depois, o vaso sanitário emite um relatório da análise de sua urina e envia uma orientação ao setor de nutrição para reajustar a sua dieta [1]. O filme se passa no ano de 2019. Estamos no ano de 2021 e, mesmo que ainda não haja uma toailete tão inteligente assim, os avanços tecnológicos dos últimos anos parecem indicar que o controle da nossa saúde pode passar por situações semelhantes.

Na época em que a *internet* era acessada através de um *modem* conectado à rede de telefonia fixa, a sua utilidade se limitava à leitura de notícias e conversas em salas de bate-papo disponibilizada pelos provedores. Alguns anos se passaram e ela passou a ser necessária para quase tudo. A sua abrangência expandiu-se de tal forma que agora as pessoas não são os únicos usuários, mas os dispositivos passaram a ter certa autonomia para enviar e receber informações pela rede mundial de computadores, o que passou a ser chamado de *Internet das Coisas* (do inglês, IoT). Somando isso com o desenvolvimento da Inteligência Artificial (IA) e dos Sistemas *Ciber* Físicos (do inglês, CPS) e aplicando o resultado aos processos de

produção, temos uma nova revolução industrial conhecida como Indústria 4.0 [2,3].

Desde 2012, quando o termo “Indústria 4.0” ganhou notoriedade [4], os países têm buscado estimular seus parques industriais a se desenvolverem nessa direção. No Brasil, o governo federal instituiu o Grupo de Trabalho para a Indústria 4.0 (GTI 4.0), do qual o Inmetro é integrante, com o objetivo de desenvolver uma agenda nacional sobre o assunto [3]. Além disso, um programa piloto do Serviço Nacional de Aprendizagem Industrial (SENAI) para implantação dos conceitos da quarta revolução industrial possibilitou um aumento médio de 22% na produtividade de micro, pequenas e médias empresas [5].

A revolução continua a se expandir não só dentro das fábricas, mas pulou os muros para alcançar outros setores, como o ambiente hospitalar, ambulatorial e de cuidados com a saúde. A Saúde 4.0, como tem sido chamada, foca na virtualização para monitorar em tempo quase real e responder a pacientes e demais envolvidos com o sistema de saúde de forma personalizada [6]. Ou seja, os dados do paciente podem ser coletados remotamente para alimentar seu prontuário eletrônico, permitindo assim a tomada de decisão médica mais precisa [7]. Outras aplicações possíveis incluem [8]:

- 
- ❑ **Telemedicina** - a realização de consultas à distância através de um smartphone ganhou projeção devido à pandemia de Covid-19. Além de mitigar contato com pessoas potencialmente portadoras de doenças infectocontagiosas, essa ferramenta reduz o tempo de espera em consultórios e permite o atendimento de pessoas em regiões remotas;
 - ❑ **Assistência domiciliar virtual** - uma espécie de “cuidador virtual” que pode suprir a necessidade de assistência médica diária de pacientes (principalmente idosos) que moram sozinhos;
 - ❑ **Rastreamento de ingestão de medicamentos** - sensores em forma de pílulas podem ser ingeridos por pacientes (por exemplo, esquizofrênicos ou bipolares) afim de lembrá-los quando e se a medicação foi tomada;
 - ❑ **Sistemas de resposta emergencial** - a monitoração constante de pacientes permite o estabelecimento mais preciso de padrões de saúde e, com isso, a intervenção proativa em caso de emergência.

Para que esse cenário venha a fazer parte do cotidiano, é necessário que o sistema de saúde atual realize algumas mudanças, entre as quais, podem ser citadas como mais críticas a utilização de tecnologias digitais não-invasivas para coletar as informações do paciente em casa (ou no local em que estiver) e de plataformas digitais para compartilhamento e tratamento das mesmas [9].

No primeiro caso, as tecnologias não-invasivas referem-se principalmente àquelas destinadas a coletar parâmetros fisiológicos como temperatura corpórea, pressão arterial, saturação de oxigênio e atividades cerebral ou cardíaca. A IoMT (*Internet das Coisas Médicas*, em tradução livre) é uma alusão à IoT e tem sido impulsionada pela miniaturização, nanotecnologia e desenvolvimento de sistemas baseados em sensores com transmissão sem fio que podem, literalmente, ser vestidos (*wearables*) [8]. No segundo caso, as plataformas digitais consistem nos processos de armazenamento de dados em nuvem e nas aplicações que os utilizam. Nesse sentido, o desenvolvimento da Inteligência Artificial e do processamento em *Big Data* permitem, por exemplo, o reconhecimento de padrões de doença como câncer e depressão [8]. Diante desse cenário, se insere a metrologia legal.

A importância da metrologia legal

Oficialmente, metrologia legal é a “prática e processo de aplicar à metrologia uma estrutura legal e regulamentadora e implementar sua execução” [10]. Por ser uma forma de regulação, ela também pode ser entendida como “uma intervenção do Estado no funcionamento da sociedade ou da economia e se dá quando a ausência de intervenção pode resultar em prejuízos ou danos, ou pode comprometer o alcance dos objetivos legítimos” [11] ou ainda como o “conjunto de instrumentos jurídico-normativos (leis, decretos, regulamentos e outras normas) de que dispõe o governo para estabelecer obrigações que devem ser cumpridas pelo setor privado, pelos cidadãos e pelo próprio governo” [12].

Em outras palavras, a metrologia legal tem a responsabilidade de garantir o funcionamento adequado dos instrumentos de medição utilizados pela sociedade tal qual, por exemplo, um departamento de manutenção tem a responsabilidade de garantir o funcionamento adequado das máquinas de uma fábrica. Contudo, a metrologia legal não se resume a um “departamento”, mas é um conjunto de organizações que, no Brasil, é chamado de Sistema Nacional de Metrologia, Normalização e Qualidade Industrial (Sinmetro) [14].

Cada integrante desse sistema executa uma parte do controle metrológico legal, que pode ser vista como um fluxo. Inicialmente, a Diretoria de Metrologia Legal do Inmetro (Dimel) elabora o Regulamento Técnico Metrológico (RTM) que define “as regras do jogo” para cada tipo de instrumento [13]. Atualmente, no âmbito da saúde, há regulamentos para esfigmomanômetros (medidores de pressão arterial) [15], termômetros clínicos [16] e balanças utilizadas, por exemplo, em farmácias de manipulação para pesagem dos componentes utilizados na fabricação dos remédios [17].

Com isso, o requerente (fabricante ou importador) que deseja comercializar um instrumento desses no Brasil deve submeter o modelo à avaliação pela Dimel, que consiste na análise de documentação e na realização de ensaios [13]. A Figura I apresenta o desempenho geral dos modelos de esfigmomanômetros avaliados nos últimos três anos levantado com base nos dados do sistema de gerenciamento de processos do Inmetro, denominado Orquestra.

Observando os dados constata-se que, em média, 46% dos esfigmomanômetros são reprovados na Avaliação de Modelo. Isso demonstra que, na ausência do controle metrológico, quase metade dos modelos comercializados no país estaria em

desacordo com as Recomendações da Organização Internacional de Metrologia Legal (OIML), já que elas geralmente são a base dos RTMs [15]. Na realidade, sem a metrologia legal, esse número poderia ser maior, pois alguns dos modelos conseguiram aprovação na segunda ou terceira tentativas.

Essa constatação ganha mais importância quando se analisa os motivos das reprovações que, no caso de esfigmomanômetros, são frequentemente relacionados a três requisitos. O primeiro refere-se àquelas situações em que o processo de Avaliação de Modelo é encerrado porque o requerente não corrige as não-conformidades dentro do prazo (reprovação por decurso de prazo). Nesse caso, é importante destacar as reprovações devido a não-conformidades encontradas nos relatórios de investigação clínica, necessários para demonstrar que o instrumento consegue realizar a medição com diferentes tipos de pacientes. Embora a exigência desse relatório e a análise de seu conteúdo tenham sido identificadas como ponto crucial para assegurar satisfatoriamente o desempenho metrológico dos esfigmomanômetros, apenas o Brasil e Cuba adotam essa prática entre os países da América Latina e Caribe [18-20].

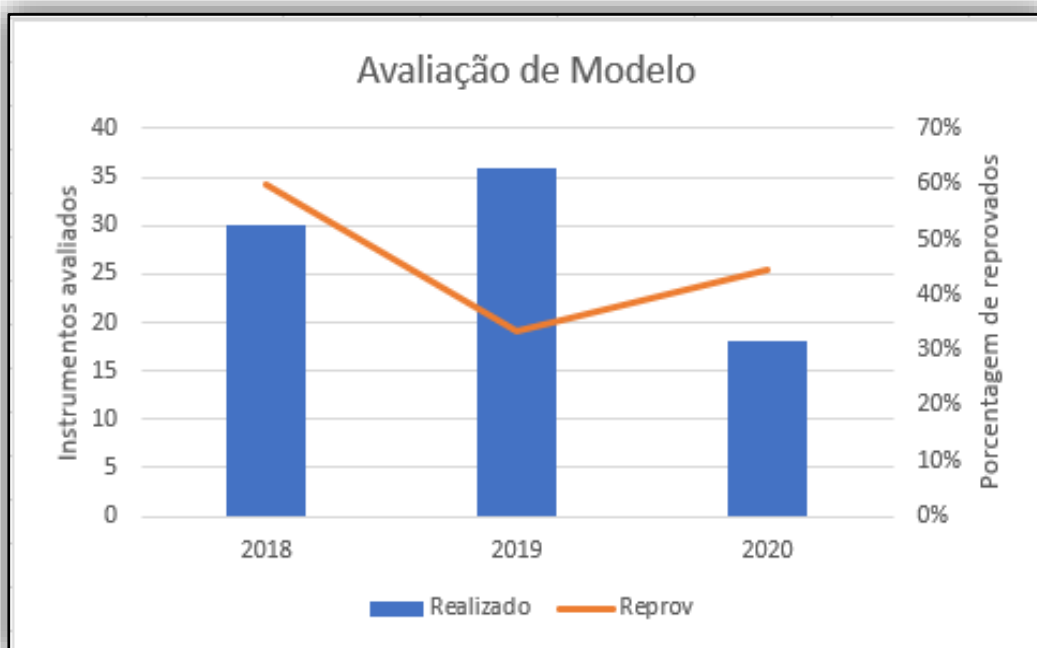


Figura I Desempenho de esfigmomanômetros na avaliação de modelo.

Quando o instrumento é aprovado em todas as etapas da Avaliação de Modelo, é emitida a Portaria de Aprovação de Modelo (PAM) atestando que o mesmo pode ser comercializado [10]. Contudo, antes de chegar ao mercado, cada exemplar fabricado ou importado deve passar pela Verificação Inicial, que pode ser vista como uma versão bem simplificada da Avaliação de Modelo [15]. Com base nos dados do Sistema de Gestão Integrada do Inmetro (SGI), a Figura 2 apresenta quantidade de esfigmomanômetros que passaram por essa etapa nos últimos três anos.



Figura 2 Quantidade de esfigmomanômetros que passaram pela verificação inicial.

Como pode ser observado, a quantidade de instrumentos verificada anualmente é grande (em torno de 1,4 milhão), indicando que o número de pessoas afetadas por um possível erro de medição seria considerável. A queda brusca de instrumentos verificados em 2019 (aproximadamente 74%) deve-se ao fato de que esses dados se referem somente às verificações realizadas pelos órgãos da Rede Brasileira de Metrologia e Qualidade do Inmetro (RBMLQ-I).

Como alternativa, as empresas podem obter autorização para declarar conformidade de seus instrumentos [22, 24] o que, de fato, aconteceu em 2019. Atualmente, há 9 empresas autorizadas e, portanto, a quantidade de esfigmomanômetros colocada anualmente no mercado é bem maior do que apresentado acima.

Outro ponto que merece destaque é a resiliência da RBMLQ-I, visto que mesmo com a limitação das atividades durante a pandemia da Covid-19, a quantidade de verificações em 2020 reduziu apenas 10% em relação ao ano anterior.

Os desafios

Depois que o esfigmomanômetro é vendido, ele deve ser verificado novamente a cada ano (Verificação Periódica) ou toda vez que passar por reparo (Verificação Eventual ou Após Reparo) em um dos órgãos da RBMLQ-I [15, 23]. A Figura 3 apresenta os dados extraídos do SGI.

Os dados apontam para uma taxa média de reprovação em torno de 7%. Caso o controle metrológico não existisse, essa proporção de I reprovação a cada 14 instrumentos teria impacto considerável na sociedade, visto que muitos desses instrumentos são utilizados em clínicas e hospitais para a medição em vários pacientes diariamente e considerando que até 10% de pessoas hipertensas não são diagnosticadas como tal caso o esfigmomanômetro apresente erro de apenas 1 mmHg acima do admissível [21]. A queda na quantidade de instrumentos avaliados em 2020 se deve à suspensão em grande parte das atividades devido à pandemia de Covid-19 [24] já que a verificação periódica é realizada nos locais em que os esfigmomanômetros são utilizados, ao contrário da verificação inicial.

Uma vez que as considerações apresentadas permitem vislumbrar a contribuição que a metrologia legal é capaz de oferecer à sociedade no âmbito da saúde, é preciso analisar os desafios que surgem com o paradigma estabelecido pela Saúde 4.0.

A primeira questão a ser analisada refere-se ao escopo da metrologia legal. Ou seja, diante do novo paradigma apresentado pela Saúde 4.0, quais tipos de instrumento de medição devem ser submetidos ao controle metrológico legal?

Como fora mencionado, os dispositivos *wearables* compõem uma parte importante dessa nova forma de gerenciar a saúde. Devido à pandemia de Covid-19, muitas pessoas compraram equipamentos desse tipo afim de monitorar sua própria saúde com o mínimo de contato social. Com isso, 700 milhões de pessoas já possuem wearables e a previsão para os próximos anos é que esse mercado movimente em torno de US\$ 70 bilhões [25, 27].

Diante desse cenário, é preciso avaliar com cautela a abrangência da metrologia legal para evitar que seus custos não sobrepujam os benefícios e nem que sua ausência desproteja a sociedade. Nesse sentido, a Análise de Impacto Regulatório (AIR) é uma ferramenta imprescindível [28] e, de certa forma, já foi iniciada com a participação do Inmetro no grupo *ad hoc* criado pela OIML para discutir o assunto.

Outra questão fundamental é quanto à forma de se realizar o controle metrológico. Uma vez que a monitoração da saúde será cada vez mais descentralizada, a localização dos instrumentos no pós-mercado também será difusa. Isso significa que as

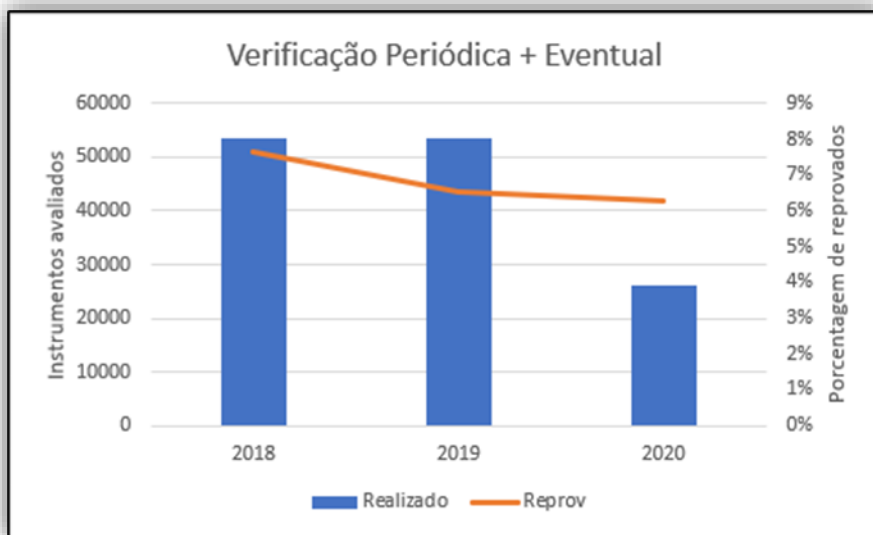


Figura 3 Dados da Verificações Periódicas e Eventuais de esfigmomanômetro.

Verificações Subsequentes (Periódicas, Eventuais e Após Reparo) necessitarão de um planejamento mais inteligente, como preconiza o atual Planejamento Estratégico do Inmetro [29].

Um caminho para se alcançar esse objetivo é a utilização de tecnologias como IoT e *blockchain* para desenvolver soluções que permitam realizar a vigilância do mercado de forma remota [26].

Outra opção é a utilização da confiabilidade estatística, que consiste em determinar a probabilidade de falha de um objeto com o passar do tempo [30]. Para ilustrar, a Figura 4 apresenta a confiabilidade de dois modelos de esfigmomanômetro a partir de alguns dados extraídos do Portal de Serviços do Inmetro nos Estados (PSIE) referentes às verificações periódicas realizadas em 2019 no Estado de São Paulo e do Espírito Santo.

As curvas exibidas indicam que, enquanto o modelo A tem 90% de chance de ser aprovado em verificação periódica se tiver até 2240 dias de utilização, o modelo B tem o mesmo desempenho com 1331 dias. Isso significa que os modelos são capazes de funcionar satisfatoriamente durante os seus 6 e 3,5 primeiros anos de vida, respectivamente. Dessa forma, as atividades poderiam ser organizadas para priorizar a verificação periódica do modelo que apresenta o menor tempo, acarretando na otimização de recursos.

Uma terceira questão refere-se aos procedimentos de ensaio. Alguns desses dispositivos *wearables* utilizam

métodos indiretos para realizar a medição (por exemplo, os termômetros clínicos de medição por radiação infravermelha e os esfigmomanômetros de medição contínua) o que impede a utilização dos padrões de trabalho atuais para determinação do erro. Assim, é necessário dispor de novos procedimentos de ensaios que sejam validados, de execução simples e de baixo custo.

E, por fim, todas essas questões tornam-se mais críticas face à restrição de recursos humanos e materiais. Nesse sentido, a busca por parcerias que viabilizem a execução de projetos é uma opção viável [29]. Como exemplo, podem ser citadas a parceria entre Dimel e o Instituto de Pesos e Medidas do Estado de São Paulo (Ipem-SP) viabilizou junto à Fundação de Amparo à Pesquisa no Estado de São Paulo (Fapesp), a implementação do primeiro programa de ensaio de proficiência registrado na base de dados do EPTIS (*European Proficiency Testing Information System*) com foco nas etapas de verificação inicial de esfigmomanômetros, possibilitando uma ferramenta de comparação interlaboratorial para as empresas que declaram conformidade de seus instrumentos [31, 32] e a parceria entre o Laboratório de Termometria (Later) da Diretoria de Metrologia Científica e Tecnologia do Inmetro (Dimci) e a Fundação Oswaldo Cruz (Fiocruz) para avaliar a exatidão dos termômetros clínicos de infravermelho e estabelecer uma correlação entre esses resultados e a calibração laboratorial.

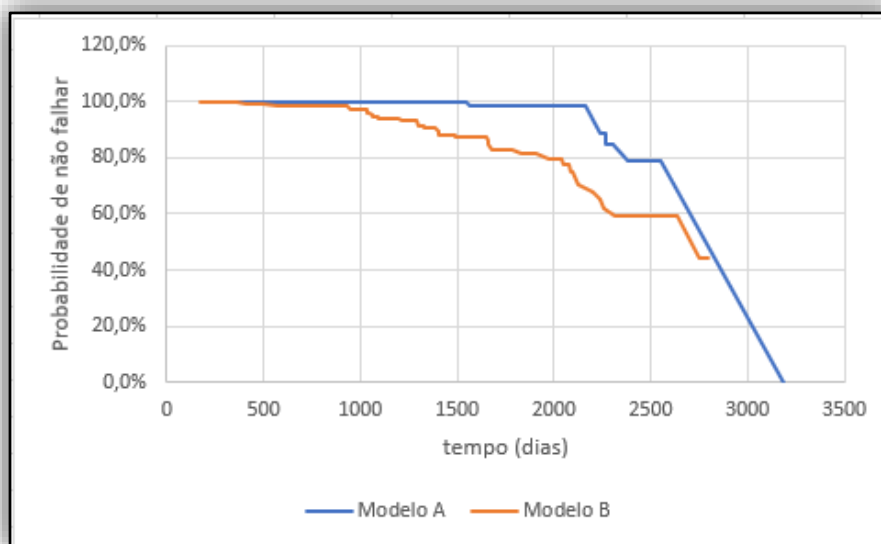


Figura 4 Curvas de confiabilidade de dois modelos de esfigmomanômetro.

Conclusão

A Saúde 4.0 apresenta um paradigma que aponta para modificar o sistema de saúde afim de torná-lo mais acessível e mais alinhado com as necessidades dos pacientes. Contudo, surgem também questões importantes que precisam ser tratadas pela metrologia legal de forma que esses mesmos pacientes continuem a ter segurança nas medições.

Referências

- [1] A ILHA. Direção de Michael Bay. Estados Unidos: Warner Bros. Pictures, 2005.
- [2] EUROPEAN COMMISSION. Digital Transformation Monitor – Germany: Industrie 4.0. Janeiro, 2017.
- [3] www.industria40.gov.br. Acessado em 29 de abril de 2021.
- [4] National Academy of Science and Engineering. Recommendations for implementing the strategic initiative Industrie 4.0. Alemanha, 2013.
- [5] <https://noticias.portaldaindustria.com.br/noticias/inovacao-e-tecnologia/empresas-ganham-em-productividade-com-a-industria-40/>. Acessado em 01/05/2021.
- [6] MONTEIRO, A. C. B; FRANÇA, R. P; ESTRELA, V. V; IANO Y. Health 4.0: Applications, Management, Technologies and Review. **Medical Technologies Journal**, Volume 2, 4ª Edição, 2018, páginas 262-276.
- [7] <https://www.gov.br/pt-br/noticias/saude-e-vigilancia-sanitaria/2020/01/governo-lanca-camara-da-saude-4-0-para-aprimorar-uso-da-tecnologia-no-setor-da-saude>. Acessado em 01/05/2021.
- [8] WORLD ECONOMIC FORUM. Health and Healthcare in the Fourth Industrial Revolution: Global Future Council on the Future of Health and Healthcare 2016-2018. Suíça, 2019.
- [9] TORTORELLA, G. L; SAURIN, T. A; FOLGIATTO, F. S; ROSA, V. M, TONETTO, L. M; MAGRABI, F. Impacts of Healthcare 4.0 digital technologies on the resilience of hospitals. **Technological Forecasting and Social Change Journal**, Volume 166, 2021.
- [10] INMETRO. Portaria n° 150, de 29 de março de 2016.
- [11] CONMETRO. Guia de boas práticas de regulamentação. 2007. Disponível em: http://www.inmetro.gov.br/qualidade/pdf/guia_p_ortugues.pdf. Acesso em: 03/05/2021.
- [12] CASA CIVIL. Diretrizes gerais e guia orientativo para elaboração de análise de impacto regulatório - AIR. Subchefia de análise e acompanhamento de políticas governamentais et al. Brasília, DF: Presidência da República, 2018. Disponível em https://www.gov.br/casacivil/pt-br/centrais-de-conteudo/downloads/diretrizes-gerais-e-guia-orientativo_final_27-09-2018.pdf/view. Acesso em: 03/05/2021.
- [13] INMETRO. Portaria n° 302, de 19 de junho de 2019.
- [14] BRASIL. Lei n° 5966, de 11 de dezembro de 1973.
- [15] INMETRO. Portaria n° 046, de 22 de janeiro de 2016.
- [16] INMETRO. Portaria n° 402, de 23 de agosto de 2019.
- [17] INMETRO. Portaria n° 236, de 22 de dezembro de 1994.
- [18] Organização Mundial da Saúde. WHO Technical Specifications for Automated Non-Invasive Blood Pressure Measuring Devices with Cuff. 2020.
- [19] SHARMAN, J. E; O'BRIEN, E; ALPERT, B *et al.* Lancet Commission on Hypertension group position statement on the global improvement of accuracy standards for devices that measure blood pressure. Janeiro, 2020. Volume 38, Issue 1, p 21-29.
- [20] LOMBARDI, C; SHARMAN, J. E; PADWAL, R *et al.* Weak and fragmented regulatory framework on the accuracy of blood pressure-measuring devices pose a major impediment for the implementation of HEARTs in the Americas. Dezembro, 2020, 22(12):2184-2191.

[21] RODRIGUES, B. A. Filho; FARIAS, R. F; ANJOS, W. E. Evaluating the impact of measurement uncertainty in blood pressure measurement on hypertension diagnosis. **Blood Pressure Monitoring**, 2018, 23: 141-147.

[22] INMETRO. Portaria n° 400, de 12 de agosto de 2013.

[23] <https://www4.inmetro.gov.br/acesso-a-informacao/institucional/localizacao/por-instituicao>. Acessado em 04/05/2021.

[24] INMETRO. Portaria n° 101, de 20 de março de 2020.

[25] <https://www.forbes.com/sites/forbestechcouncil/2021/01/15/welcome-to-the-human-data-economy/?sh=388343df76eb>. Acessado em 04/05/2021.

[26] MELO, W. S. Jr; TARELHO, L. V. G; RODRIGUES, B. A. Filho; BESSANI, A. N; CARMO, L. F. R. C. Field surveillance of fuel dispensers using IoT metering and blockchains. **Journal of Network and Computer Applications**. Volume 175, Fevereiro, 2021.

[27] <https://www.cnbc.com/2020/11/20/samsung-wearable-device-sales-are-up-more-than-30percent-this-year.html>. Acessado em 04/05/2021.

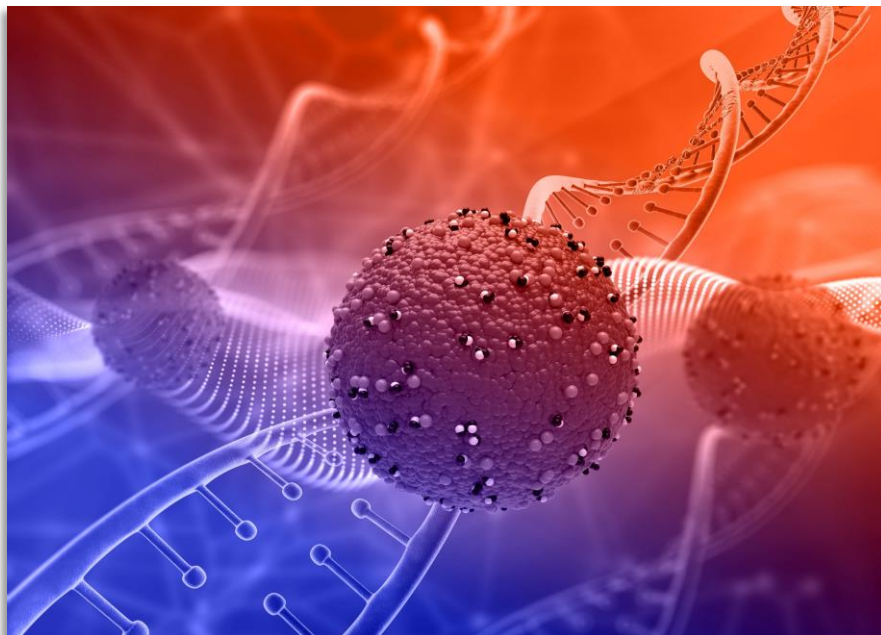
[28] BRASIL. Decreto n° 10411, de 30 de junho de 2020.

[29] INMETRO. Planejamento Estratégico 2021-2023. Brasília, DF. 2021.

[30] <http://www.portalaction.com.br/confiabilidade>. Acessado em 04/05/2021.

[31] RODRIGUES, B. A. Filho; FARIAS, R. F; ANJOS, W. E. Implementing a proficiency test provider for sphygmomanometers in Brazil. **Journal of Physics Conference Series**. Junho, 2018.

[32] RODRIGUES, B. A. Filho; FARIAS, R. F; ANJOS, W. E. Evaluating the Performance of Laboratories Testing Sphygmomanometers in Legal Metrology due to Proficiency Testing. **Journal of Metrology Society of India**. Agosto, 2020.



Desafios da contagem de células humanas na biometrologia

Daniella Paiva Bonfim

Diretoria de Metrologia Aplicada às Ciências da Vida (Dimav)

Fernanda Leve

Diretoria de Metrologia Aplicada às Ciências da Vida (Dimav)

E-mail: fleve@Inmetro.gov.br

As células são as unidades básicas dos organismos vivos e podem apresentar diferentes complexidades. Seu uso em diagnóstico clínico iniciou os estudos em Biologia Celular, mas o avanço da Biotecnologia e a inserção das células na indústria trouxeram a demanda de validar métodos de análise de células bem como de harmonizar protocolos e processos em todo o mundo. Cabe mencionar que a análise de células inclui medições de componentes intra e extracelulares (por exemplo, elementos da matriz extracelular; produtos secretados por células como enzimas, anticorpos, etc.); interações célula-célula e célula-matriz; etc.), e medições das células em si, que podem estar relacionadas ao tamanho, número de células em determinada população, função biológica e quantificação de componentes celulares (DNA, proteína e lipídeos); no entanto, diversos fatores dificultam a realização dessas medições, trazendo a necessidade de desenvolver métodos precisos e com fundamentos metroológicos nessa área, que permanecem escassos. Nesse sentido, alguns esforços vêm sendo realizados para criar métodos de garantia de medição de células e melhorar a reprodutibilidade e a rastreabilidade dos resultados.

Breve histórico dos métodos de contagem de células

As primeiras técnicas para contagem de células sanguíneas foram desenvolvidas no século XIX, com o objetivo de investigação e diagnóstico de doenças. Em 1852, o professor Karl Vierordt realizou as primeiras contagens dessas células em um capilar de vidro de diâmetro conhecido, cuja capacidade poderia ser medida: o sangue era adicionado em uma lâmina coberta com uma fina camada de albumina e a contagem de células sanguíneas era realizada em microscópio [1]. No entanto, após alguns estudos, Georges Hayem, em 1875, desenvolveu os primeiros hemocitômetros, instrumentos onde uma gota de sangue era aplicada diretamente em um orifício no centro de uma placa de vidro e coberta com uma lamínula. As células eram então visualizadas, no microscópio, em uma única grade de contagem contendo 16 quadrantes separados por linhas simples. Porém, como a câmara não era preenchida por ação capilar, a distribuição homogênea das células era um problema e fez com que o dispositivo de Hayem não tenha sido bem aceito na época [2].

Diversas melhorias foram realizadas no hemocítômetro por outros pesquisadores ao longo do tempo, incluindo a criação de dois lados para a contagem, o uso da ação capilar para preencher a câmara e modificações nas grades de contagem, até que O. Neubauer simplificou o *layout* da câmara em 1922, resultando na câmara de Neubauer, rotineiramente utilizada até os dias atuais para contar outros tipos de células, além das sanguíneas [3]. A câmara de Neubauer consiste em uma lâmina grossa de vidro com duas câmaras independentes, uma superior e uma inferior (Fig. 1). A grade de contagem tem 3 mm x 3 mm de tamanho, sendo subdividida em nove quadrantes de 1 mm x 1 mm. Para contagem, uma lamínula é colocada sobre a câmara, cobrindo a área central, onde se encontra a grade de contagem. Dessa forma, quando uma amostra contendo células é coberta com a lamínula, a ação capilar preenche completamente a câmara com a amostra [4]. A contagem manual em câmara de Neubauer ainda é considerada um método padrão-ouro [5], mas mesmo com baixo custo, é um método trabalhoso, demorado e que necessita de treinamento [6].

Com o desenvolvimento e crescimento da indústria biotecnológica mundial, a quantificação celular deixou de ser apenas utilizada em diagnóstico clínico e na pesquisa em ensaios de toxicidade e viabilidade celular, e passou a ser de fundamental importância também na fabricação de produtos de alto valor agregado. Isso inclui produtos biológicos¹ e células manipuladas para uso em terapia celular². O mercado global de produtos biológicos, mais especificamente anticorpos monoclonais, foi avaliado em aproximadamente US\$ 115,2 bilhões e deve gerar receita de US\$ 300 bilhões até 2025 [8]. Já o mercado global de terapia celular é impulsionado, principalmente, pelo crescente número de estudos clínicos em terapias com células-tronco e foi estimado em US\$ 7,8 bilhões em 2020, com previsão de chegar a US\$ 9,5 bilhões em 2021, mostrando que esse segmento segue em crescimento constante [9].

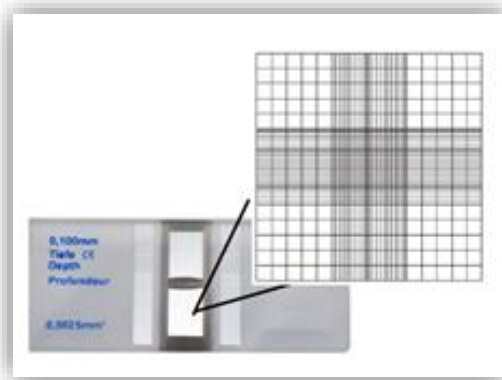


Figura 1 Câmara de Neubauer. A contagem de células por essa metodologia é geralmente realizada pela quantificação nos quatro quadrantes mais extremos: após a contagem, o cálculo do número total de células é realizado através da fórmula “ $N = ((Q1+Q2+Q3+Q4)/4) \times 10^4$ ” e, dessa forma, obtém-se o número total de células/mL. Em maior aumento, pode-se observar a grade de contagem da câmara de Neubauer.

¹ Produto biológico: medicamento biológico não novo ou conhecido que contém molécula com atividade biológica conhecida, já registrado no Brasil e que tenha passado por todas as etapas de fabricação (formulação, envase, liofilização, rotulagem, embalagem, armazenamento, controle de qualidade e liberação do lote de produto para uso) [7].

² Produtos para terapia celular: células de origem autóloga ou alogênica, em que as células foram propagadas, expandidas, selecionadas, tratadas farmacologicamente ou de outra forma alteradas nas características biológicas *ex vivo*, para administração em pacientes e que é aplicável à prevenção, tratamento, cura, diagnóstico ou mitigação de doenças ou ferimentos [14].

Dessa forma, métodos automáticos de contagem foram desenvolvidos com o objetivo de fornecer a quantidade de células em menor tempo e com maior rendimento, precisão e reprodutibilidade [10]. Em geral, os instrumentos automatizados de contagem de células consistem em uma câmara digital para a obtenção de imagens e as análises são realizadas por meio de um *software* especializado que requer um envolvimento mínimo do usuário [11]. Existem vários sistemas automatizados de contagem de células disponíveis (Fig. 2), incluindo *Cedex HiRes Analyzer System* (Roche), *Luna™ Automated Cell Counter* (Logos Biosystem), *TC10™ e TC20™ Cell Counters* (Bio-Rad) e *Countess® Automated Cell Counter* (Invitrogen) (Fig. 2) [6].

Dentre os métodos automatizados de contagem de células, o mais amplamente utilizado é a citometria de fluxo (Fig. 3) baseada em fluorescência, que surgiu na década de 1960. Resumidamente, o método se baseia na passagem de células individuais por um fluxo contínuo, por onde incide um *laser* e, através do espalhamento da luz incidente, é possível distinguir parâmetros morfológicos de tamanho (*forward scattering*, FSC) e granulosidade (*side scattering*, SSC), que representam a complexidade interna dessas células [12]. Existem diferentes tipos de citômetros de fluxo para diferentes aplicações. Para contagem

absoluta do número de células, dependendo do modelo utilizado, é necessário utilizar padrões de referência, isto é, um material que contém uma concentração precisa e pré-estabelecida de partículas ou células fluorescentes de um tamanho definido [13]. Dentre os desafios da contagem de células com precisão, está a ausência de materiais de referência³ para esta finalidade.

Por que ainda há desafios na contagem de células?

Conforme mencionado, a demanda crescente por serviços e produtos biotecnológicos evidenciou a necessidade de desenvolver métodos para uma medição mais precisa de células, uma vez que erros no processo de contagem podem gerar resultados não confiáveis e não reprodutíveis. A falta de reprodutibilidade de dados na pesquisa científica interfere no desenvolvimento de diferentes setores da biotecnologia e tem impactos negativos sobre a saúde, menor eficiência da produção e desperdício de tempo e de dinheiro [16]. Erros de identificação de células, por exemplo, contribuem fortemente para a irreprodutibilidade da pesquisa científica e estima-se que apenas 19% dos trabalhos publicados realizaram testes de autenticação, isto é, certificação de origem, nas linhagens celulares utilizadas [17]. Nesse contexto,



Figura 2 Contador Automático *Countess®* (Invitrogen), que fornece o número total de células/mL, incluindo dados da viabilidade celular, permitindo a observação das células em seu monitor (maior aumento).

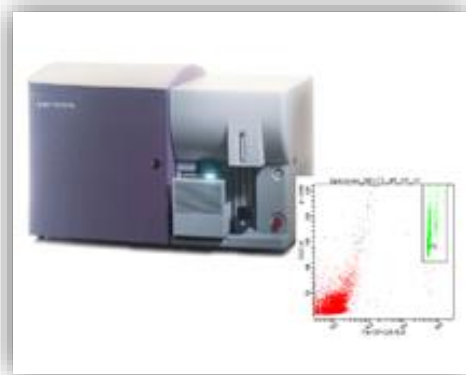


Figura 3 Citômetro de fluxo (BD Biosciences) e gráfico ilustrativo mostrando a população de células fluorescentes (em vermelho) e a população de padrões de referência (em verde), obtidas por essa metodologia.

³ *Materiais de referência: materiais suficientemente homogêneos e estáveis em relação a propriedades específicas, preparados para se adequar a um uso pretendido [15].*

o surgimento de bancos repositórios de células em todo o mundo teve forte contribuição na melhoria da qualidade das células distribuídas pelos laboratórios de pesquisa e da indústria, sendo importantes na harmonização de protocolos, na definição de características fenotípicas e genotípicas de cada linhagem celular.

Porém, em relação à contagem de células, a dificuldade de estabelecer métodos reprodutíveis também surge pela complexidade inerente de organismos vivos e seus componentes, incluindo as células em si: mesmo células de uma única população apresentam diferenças em seu tamanho; no seu metabolismo; em sua complexidade interna, por exemplo, no conteúdo de grânulos, vesículas e organelas; podem estar em diferentes fases no ciclo celular; e todas essas características podem interferir no resultado da contagem, dependendo do método utilizado. Ainda, a homogeneização extensiva pode danificar a célula e até induzir um processo de morte celular, alterando sua morfologia e, conseqüentemente, a contagem, ou ainda

comprometer as etapas de diluições necessárias no processo; o tempo entre homogeneizações também pode interferir no número de células amostral que será contado em dado experimento [18]. Outros fatores que interferem no processo de contagem estão ilustrados na Figura 4: a manutenção inadequada das linhagens celulares; a seleção de reagentes inadequados; o uso de instrumentos de medição inadequados ou não calibrados; a falta de protocolos harmonizados; a falta de treinamento do operador/analista; erros na aquisição e na análise de dados; além da dificuldade em realizar muitas repetições nas medições em decorrência do alto custo [19]. Em conjunto, esses fatores representam fontes potenciais de variabilidade e podem influenciar na reprodutibilidade dos dados obtidos.

Também, existe a dificuldade de desenvolver materiais de referência baseados em células eucarióticas, principalmente devido à variabilidade inerente aos materiais biológicos [20].

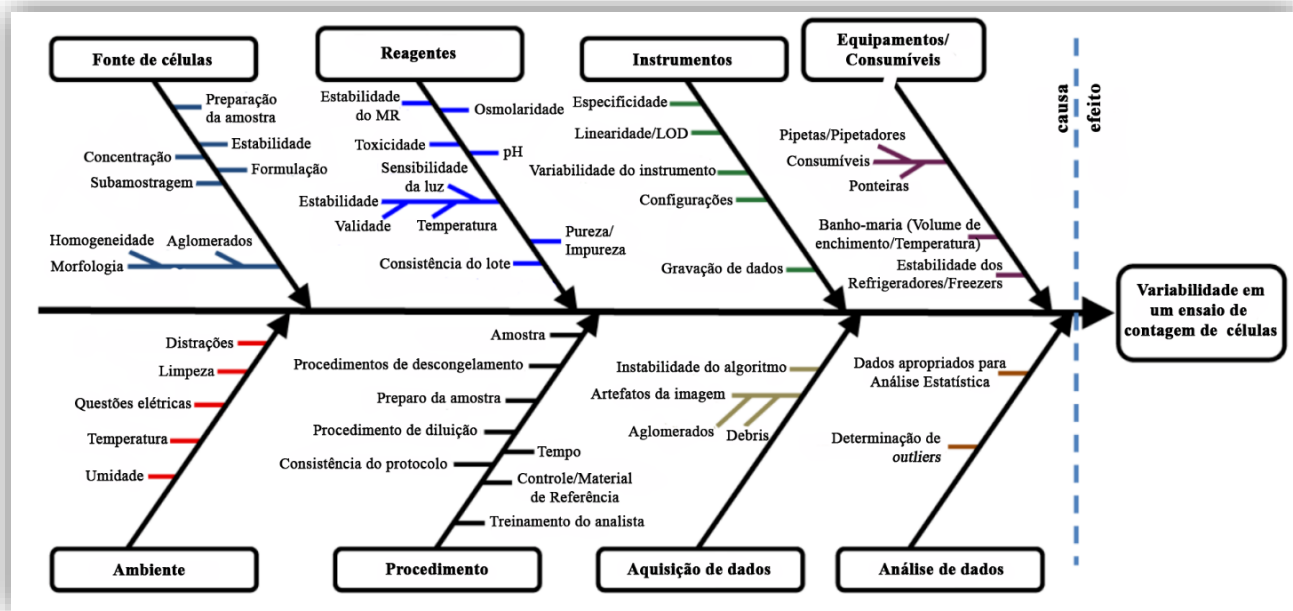


Figura 4 Exemplo de diagrama *Ishikawa*, considerando potenciais fontes de variabilidade em ensaios de contagem de células (Adaptado de [19]).

⁴ *Reprodutibilidade: precisão de medição (concordância entre valores medidos) conforme um conjunto de condições de reprodutibilidade (medições de um mesmo objeto ou de objetos similares, realizadas em locais diferentes, por diferentes operadores)*[15].

Como superar os desafios na contagem de células?

Algumas iniciativas têm sido tomadas com o objetivo de estabelecer métodos rastreáveis e reprodutíveis de contagem de células. Em 2014, foi criado um grupo de trabalho em análises de células denominado *Cell Analysis Working Group* (CAWG), que faz parte do *Consultative Committee for Amount of Substance* (CCQM), um comitê consultivo para metrologia química e biológica, do *Bureau International des Poids et Mesures* (BIPM). O CAWG possui 26 membros e tem a missão de identificar, estabelecer e sustentar a comparabilidade global das capacidades de medição de células, através de sistemas de medição de referência de ordem metrológica, com rastreabilidade para o Sistema Internacional de Medidas (SI) ou outras unidades acordadas [21]. Ainda recente, o CAWG realizou apenas três estudos piloto entre os institutos de metrologia na área de células humanas: o P123, de quantificação de células aderidas a um substrato sólido, que teve a participação do INMETRO; o P165, de quantificação de células CD34+ em citômetro de fluxo [22]; e o P217, em andamento e também com participação do INMETRO, para quantificação de células sanguíneas por diferentes métodos. Até o momento ainda não foi realizada nenhuma comparação-chave nessa área, mas a criação de um grupo específico na análise de células dentro do CCQM, mesmo que com estudos ainda preliminares, demonstra o crescimento desse tema no mundo.

Outra iniciativa relevante nesse setor foi a publicação da norma ISO 20391, um guia geral nos métodos de contagem de células. Tal norma se divide em duas partes, onde a primeira, publicada em 2018 [23], define os termos de contagem de células; classifica e descreve os métodos para células em suspensão, nos quais a contagem é expressa como concentração celular, ou seja, número de células por volume; e para células aderidas, expressas como densidade de células, ou seja, número de células por unidade de área. Nesse contexto, cabe ressaltar a dificuldade em adotar uma unidade internacional de medidas que expresse a quantificação de células, permitindo a sua rastreabilidade ao SI. No entanto, e com a recente redefinição do SI, mais especificamente da redefinição do conceito do mol como a unidade para quantidade de matéria, isto é, a medida do número

de entidades elementares especificadas, e considerando que o termo “entidades elementares” se refere a quaisquer objetos ou eventos contáveis, tal dificuldade agora pode ser superada, permitindo que células, bactérias, vírus e outros organismos vivos podem ser considerados como entidades elementares, dependendo da capacidade de determiná-los por meio da contagem [24, 25].

Já a segunda parte da ISO 20391 [26], publicada em 2019, dá orientações em relação aos procedimentos de análise estatística de dados e avaliação da qualidade de uma medição de contagem de células. Esse documento fornece orientações para a seleção do desenho experimental apropriado, considerando o número de diluições, número de replicatas e números de medições. Além disso, em relação a análise de dados, e na ausência de um material de referência, a qualidade de uma medição de contagem de células pode ser avaliada por meio de indicadores de repetibilidade e de proporcionalidade. Por sua vez, o princípio de proporcionalidade fornece um modelo idealizado que permite avaliar um aspecto crítico do desempenho de um método de contagem de células. Dessa forma, a implementação da análise de dados de contagem de células, de acordo com a ISO 20391-2, representa uma importante etapa para tornar os métodos de contagem reprodutíveis.

Em conclusão, embora a contagem de células seja um processo aparentemente simples, bem estabelecido, e realizado há muitos anos, apenas com os avanços da indústria biotecnológica e necessidade de contagem de células em larga escala, surgiram os métodos automatizados. No entanto, a padronização de métodos e o desenvolvimento e adoção de estratégias de garantia de medição ainda requerem muitos estudos e apenas passos iniciais nesse tema de grande impacto na indústria biotecnológica e da saúde foram dados.

Referências

- [1] MATH, V. M.; et al. Red Blood Cell Count: Brief History and New Method. **MGM Journal of Medical Sciences**. 3(3): 116-119
- [2] VERSO, M. L. The evolution of blood-counting techniques. **Med. Hist.** 8(2):149-158, 1964
- [3] DAVIS, J. D. The hemocytometer and its impact on progressive-era medicine. In: Department of History, University of Illinois at Urbana-Champaign, 1995
- [4] LONZA. Cell Counting and Determination of Viability via Hemocytometer. Technical Reference Guide. 2009
- [5] ZHANG, M.; et al. Improvement of cell counting method for Neubauer counting chamber. **Journal of Clinical Laboratory Analysis**. 34(1): e23024, 2019
- [6] CADENA-HERRERA, D.; et al. Validation of three viable-cell counting methods: Manual, semi-automated and automated. **Biotechnology Reports**. 7(2015): 9-16, 2015
- [7] RDC N° 55 ANVISA, de 16/12/2010
- [8] LU, R.; et al. Development of therapeutic antibodies for the treatment of diseases. **Journal of Biomedical Science**. 27:1, 2020
- [9] GRAND VIEW RESEARCH. Cell Therapy Market Size, Share & Trends Analysis Report By Use-type, By Therapy Type (Autologous, Allogenic), By Region (North America, Europe, Asia Pacific, Latin America, MEA), And Segment Forecasts, 2021-2028. **Market Analysis Report**, 2021
- [10] LOUIS, K. S. & SIEGEL, A. C. Cell viability analysis using trypan blue: manual and automated methods. **Methods Mol Biol**. 740: 7-12, 2011. doi: 10.1007/978-1-61779-108-6_2
- [11] THOLUDUR, A.; et al. Comparing Automated and Manual Cell Counts for Cell Culture Applications. **Bioprocess Technical**. 28-34, 2006
- [12] VEMBADI, A.; et al. Cell Cytometry: Review and Perspective on Biotechnological Advances. **Front. Bioeng. Biotechnol**. 7: 147, 2019
- [13] VERBRUGGHE, K.; et al. Determining Cell Concentration by Direct Volume on the BD Accuri™ C6. BD Biosciences, 2012
- [14] ISO 13022:2012 – Medical products containing viable human cells – Application of risk management and requirements for processing practices
- [15] VIM. Vocabulário Internacional de Metrologia: Conceitos Fundamentais e Gerais e Termos Associados. Duque de Caxias, RJ: INMETRO, 2012
- [16] ATCC. Disponível: <https://www.nature.com/articles/d42473-019-00004-y>. Acesso em 27/04/21
- [17] NATURE. An update on data reporting standards. *Nature cell biology*. 16(5): 385, 2014
- [18] HIRSCH, C. & SCHILDKNECHT, S. In Vitro Research Reproducibility: Keeping Up High Standards. **Frontiers in Pharmacology**. 10:1484, 2019
- [19] SIMON, C. G.; et al. Strategies for Achieving Measurement Assurance for Cell Therapy Products. **Stem Cell Translation Medicine**. 5(6): 705-708, 2016
- [20] FARUQUI, N.; et al. Cellular Metrology: Scoping for a Value Proposition in Extra- and Intracellular Measurements. **Front. Bioeng. Biotechnol**. 7: 456. 2020
- [21] BIPM. Disponível em <https://www.bipm.org/en/committees/cc/ccqm/>. Acesso em 28/04/21
- [22] CCQM. <https://www.bipm.org/documents/20126/2071059/CCQM+Strategy.pdf/31283069-94f4-f2c7-bbfc-7d652c9b3de8>. Acesso em 04/05/2021
- [23] ISO/IEC 20391-1:2018 – General Guidance on cell counting methods
- [24] THE INTERNATIONAL SYSTEM OF UNITS (SI). Bureau International des Poids et Mesures. 9th edition. 2019
- [25] GARRIDO, B. C. A Redefinição do Mol: Avanços nas Medições em Química e Biologia. In: **Cadernos de Metrologia – Especial SI/2019**, p. 41-42, 2019
- [26] ISO/IEC 20391-2:2019 – Experimental design and statistical analysis to quantify counting method performance

Princípios das Boas Práticas de Laboratório (BPL) e o Grau de Percepção por Estudantes e Pesquisadores da Biotecnologia



Elisa Rosa dos Santos

Diretoria de Metrologia Aplicada às Ciências da Vida (Dimav)

E-mail: erosa@inmetro.gov.br

José Mauro Granjeiro

Diretoria de Metrologia Aplicada às Ciências da Vida (Dimav)

E-mail: jmgranjeiro@inmetro.gov.br

Introdução

A Política de Desenvolvimento da Biotecnologia no Brasil, estabelecida no Decreto no 6.041, de 8 de fevereiro de 2007 descreve, entre outros, os objetivos específicos da ação estruturante “*Infraestrutura*”, que compreendem a criação de uma rede de instituições e empresas, públicas e privadas, com o foco em bioensaios, testes clínicos e pré-clínicos, adequada às normas internacionais de certificação de qualidade; a indução e fortalecimento dos centros de excelência em testes clínicos, pré-clínicos e bioensaios nas diversas regiões do território nacional, em conformidade com as normas internacionais de certificação de qualidade, com vistas a estimular o desenvolvimento regional nas áreas de apoio da bioindústria; e a identificação e o fortalecimento de laboratórios nacionais estratégicos de P&D capacitados a atender a demanda por produtos e processos inovadores para fortalecimento da bioindústria brasileira [1].

No que se refere à criação de redes em função da Política Brasileira de Biotecnologia (PBB), vale destacar o surgimento nos últimos anos das redes biotecnológicas: Renorbio (Rede Nordeste de Biotecnologia); BioNorte (Rede de Biodiversidade e Biotecnologia da Amazônia Legal); Pró-Centro Oeste (Rede Centro-Oeste de Pós-graduação, Pesquisa e Inovação); SulBiotec (Associação de Biotecnologia da Região Sul); e Biotecmar (Rede Nacional de Pesquisa em Biotecnologia Marinha). Com foco no

desenvolvimento regionalizado, as redes envolvem instituições atuantes na pesquisa, desenvolvimento e inovação, considerando as características locais e regionais [2-6].

Evidencia-se que a formação dessas redes foi de fundamental importância, uma vez que a biotecnologia é caracterizada pela sua multidisciplinaridade e complexidade de conhecimento, por isso seu desenvolvimento está totalmente relacionado com o estabelecimento de parcerias interorganizacionais e com a formação de redes de cooperação, sendo que estas se apresentam cada vez mais densas e compostas de uma grande diversidade de atores [7].

Quanto à PBB e a interação da bioindústria com centros geradores de conhecimento, verifica-se que essa ligação é extremamente importante, uma vez que a biotecnologia é desenvolvida por meio de processos e modelos complexos com uso de metodologias e técnicas envolvendo proteínas e outras moléculas, cultura e engenharia de células e tecidos, vetores de genes e RNA, bioinformática, nanobiotecnologia, entre outros [8].

Nesse contexto, considera-se necessário que as atividades biotecnológicas sejam realizadas por profissionais altamente qualificados como graduados, mestres, doutores e pós-doutores,

fazendo com que o mercado das empresas em biotecnologia tenha uma relação direta com universidades e institutos de pesquisa. O viés positivo dessa cooperação é que tais empresas de biotecnologia podem estar na fronteira científica, acessando conhecimentos específicos e acelerando e viabilizando seus processos de inovação [9].

Sobressai o fato de que a inovação biotecnológica está diretamente ligada à formação de recursos humanos com competência em diversas áreas do conhecimento, destacando a importância das universidades e centros de pesquisa no ensino, no treinamento de pessoal, na pesquisa e na expansão da base de conhecimento e no do empreendedorismo acadêmico. O capital humano é certamente um fator essencial para o sucesso do esforço de inovação, auxiliando no estímulo de massa crítica de profissionais de biotecnologia e áreas afins para resolver problemas e estabelecer uma base científica e produção de conhecimento [10,11]

No tocante a PBB e o fortalecimento dos centros de excelência em testes pré-clínicos e bioensaios em conformidade com as normas internacionais de qualidade, com vistas a estimular o desenvolvimento regional nas áreas de apoio da bioindústria, é fato que a implantação do sistema de gestão da qualidade provê a modernização de metodologias mais exatas e abrangentes, assim como a utilização de equipamentos modernos e sofisticados, implicando em aumento da capacitação de profissionais e demandando uma atualização contínua para exercer sua profissão com a qualidade necessária às exigências dos clientes e da coletividade [12,13].

Por sua característica complexa e tecnológica, verifica-se que a implementação de um sistema de gestão da qualidade em estudos baseados em processos e produtos biotecnológicos é fundamental para garantir transparência, interpretação correta, análise e conclusão segura de testes, documentação detalhada do trabalho de laboratório e atribuição explícita da responsabilidade pela execução de várias etapas de um experimento [14]. Assim, espera-se contribuir para a redução dos impactos da crise de reprodutibilidade no setor.

Quando se trata de qualidade em estudos pré-clínicos, existe hoje grande preocupação pelo fato destes impactarem diretamente no desenvolvimento de novas moléculas para as mais diversas finalidades, incluindo produtos da biotecnologia, que afetam a saúde das pessoas e do ambiente, com um alto custo financeiro [15,16].

No mundo, incluindo o Brasil, os estudos regulatórios pré-clínicos devem ser executados em instalações de testes segundo os princípios das Boas Práticas de Laboratório (BPL), que são a base de um sistema da qualidade laboratorial, abrangendo o processo organizacional e as condições em que esses tipos de estudos são planejados, gerenciados, desenvolvidos, monitorados, registrados, arquivados e relatados [17-19].

A união desses princípios e sua implementação visam garantir qualidade, confiabilidade, rastreabilidade e reprodutibilidade de dados laboratoriais, conforme verificado na Figura I.



Figura I Os princípios das BPL: reprodutibilidade, qualidade, confiabilidade e integridade dos dados.

Ressalta-se que, no Brasil, o Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO), realiza o reconhecimento da conformidade aos princípios das BPL, por meio da confirmação do nível de aderência da instalação teste aos Princípios das BPL e a inclusão da mesma no Programa Brasileiro de Monitoramento da Conformidade aos Princípios das BPL por meio de inspeção e de auditorias de estudos *in loco* [20].

A disseminação do conhecimento sobre sistema da qualidade Princípios das BPL, por meio de treinamento específico, é essencial para a formação de pessoal, com foco na compreensão dos princípios, seus objetivos e sua implementação pois auxilia na implementação dos requisitos da qualidade nos laboratórios, especialmente quando se envolve o mais alto nível de gerenciamento da instituição, facilitando a implementação em todos os níveis da organização. No entanto, o treinamento em BPL não é um “evento único”, é um processo periódico e todas as providências necessárias devem ser tomadas para o treinamento frequente, a fim de atualizar os profissionais e pesquisadores de laboratório, porque o Ser Humano, por sua própria natureza, é propenso a erros [21].

Destarte, essa pesquisa visou mapear o grau de percepção dos estudantes de cursos de biotecnologia do país sobre o sistema de qualidade Princípios das Boas Práticas de Laboratório (BPL), para que seu resultado possa apoiar medidas específicas relacionadas à disponibilidade de treinamentos nesse tema para um público heterogêneo como o segmento da biotecnologia.

Metodologia

A pesquisa foi planejada e realizada tendo como ponto de partida a necessidade de mapear o grau da percepção sobre os Princípios das Boas Práticas de Laboratório, considerando uma amostragem do segmento da biotecnologia. Os atores agidos são aqueles que fazem parte da coleta de dados e que, no caso desta pesquisa, inclui uma amostra de alunos de cursos de pós-graduação em biotecnologia do país.

O questionário foi elaborado de forma que pudesse realizar coleta de dados sobre a localização, a área de atuação, tipo de sistema da qualidade implementado, existência ou não de disciplina relacionada a sistema de qualidade laboratorial durante a formação, bem

como dados demográficos do respondente como idade e titulação. O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de Pelotas (número do parecer: 3.254.478) e os participantes da pesquisa tiveram acesso ao questionário somente após leitura do Termo de Consentimento Livre e Esclarecido, com a opção do respondente aceitar ou não participar da pesquisa.

Para fins de delimitar a pesquisa e para que a aplicação do questionário pudesse alcançar uma amostra representativa de alunos de pós-graduação em biotecnologia do país, foram realizados contatos telefônicos com os Coordenadores das Redes de Biotecnologia e Programas de Pós-graduação das diversas regiões, a fim de solicitar auxílio na divulgação da pesquisa.

Como estratégia complementar para atingir essa amostra populacional os autores da pesquisa e deste artigo realizaram palestras no modo à distância sobre “Reprodutibilidade de Dados e os Princípios das BPL” e, ao final, disponibilizaram um *QRCode* para que os participantes pudessem ter acesso ao questionário e emitirem suas respostas via equipamento de celular, antes mesmo de retirarem-se do ambiente da videoconferência.

Para favorecer o envio e o recebimento das respostas, o questionário foi convertido para a versão do formulário online gratuito “*Google forms*” e elaborado de forma que pudesse realizar coleta de dados sobre a instalação de teste/laboratório e dados demográficos do respondente.

Resultados

No que se refere à estratégia de obter o maior número possível de respondentes, considerou-se que a mesma obteve sucesso uma vez que houve um retorno de respondentes ao questionário de quase 100% daqueles alunos que participaram das palestras por videoconferência, com alcance em instituições de ensino e pesquisa das cinco regiões do país: norte (Amazonas), nordeste (Rio Grande do Norte e Sergipe), centro-oeste (Goiás), sudoeste (Espírito Santo e Rio de Janeiro) e Sul (Rio Grande do Sul).

Cento e vinte e um (121) foi o número total de respondentes do questionário, o qual iniciou com pergunta sobre a localização do laboratório em relação à Unidade da Federação.

De acordo como Gráfico I (Fig. 2), o maior número de respondentes faz parte de laboratórios localizados no Rio de Janeiro, porém houve uma distribuição representativa de participantes de outros estados englobando todas as regiões geográficas do Brasil, ou seja, respondentes localizados na AM (Norte); RN e SE (Nordeste); GO (Centro-Oeste); ES e RJ (Sudeste) e RS (Sul).

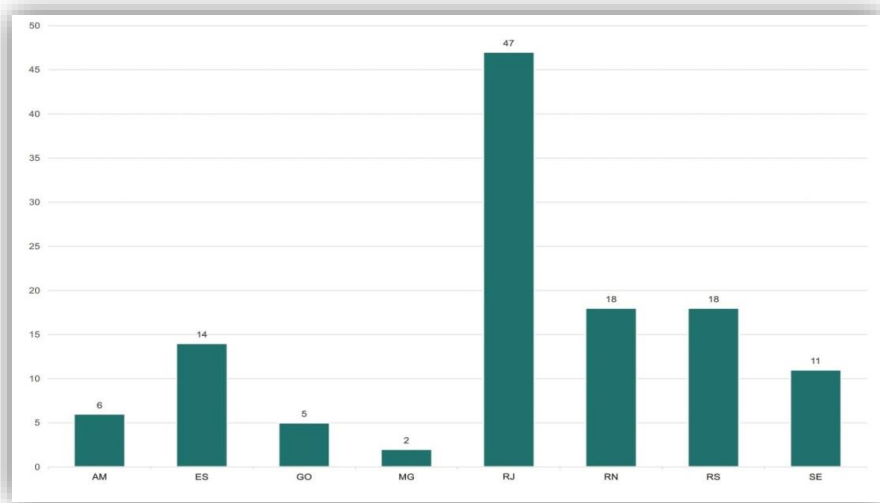


Figura 2 Localização do laboratório segundo a unidade da federação.

A segunda questão foi sobre qual instituição ou Rede a que o laboratório pertence. O Gráfico 2 (Fig. 3) explicita o número de respondentes por cada instituição e pelas redes Bionorte e Renorbio.

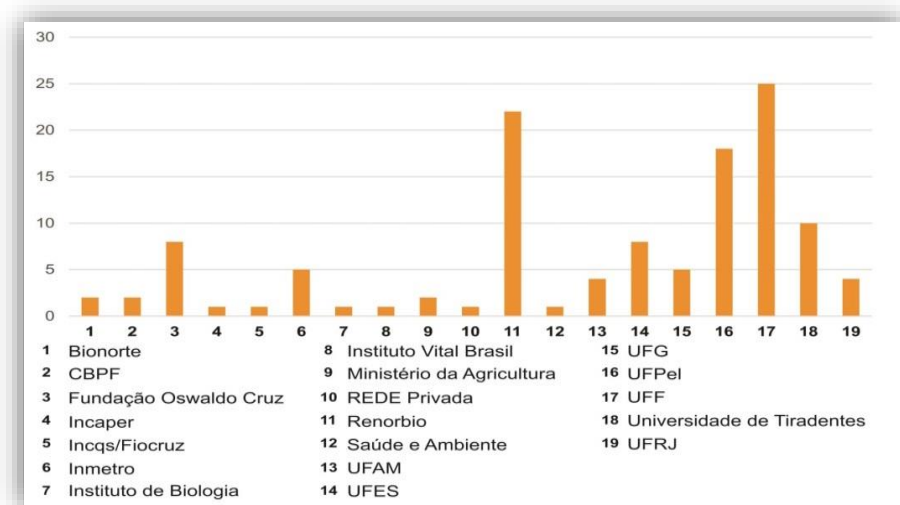


Figura 3 Número de respondentes por instituição e pela rede de biotecnologia.

No que se refere à área de atuação do Laboratório, Gráfico 3 (Fig. 4), de acordo com a classificação utilizada, a maioria (65%) respondeu que atua na área de Biotecnologia em Saúde, seguida da área Biotecnologia em Agropecuária (14%), Biotecnologia em Recursos Naturais (12%) e por último Biotecnologia Industrial (9%).

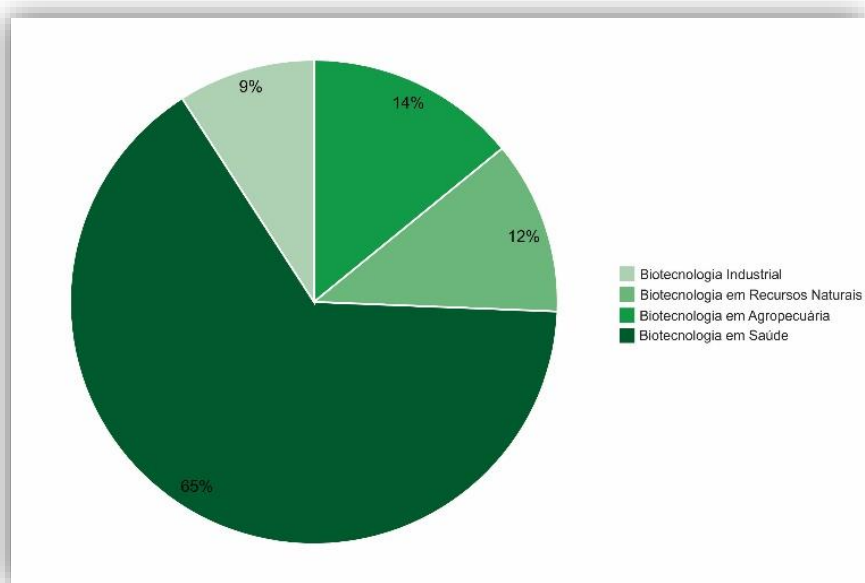


Figura 4 Área de atuação do laboratório.

Quanto a possuir ou não algum tipo de sistema da qualidade implementado no laboratório que atua, de acordo com o Gráfico 4 (Fig. 5), 78% dos respondentes afirmaram não possuir qualquer tipo de sistema de qualidade e 22% afirmaram possuir algum tipo de sistema de qualidade implementado.

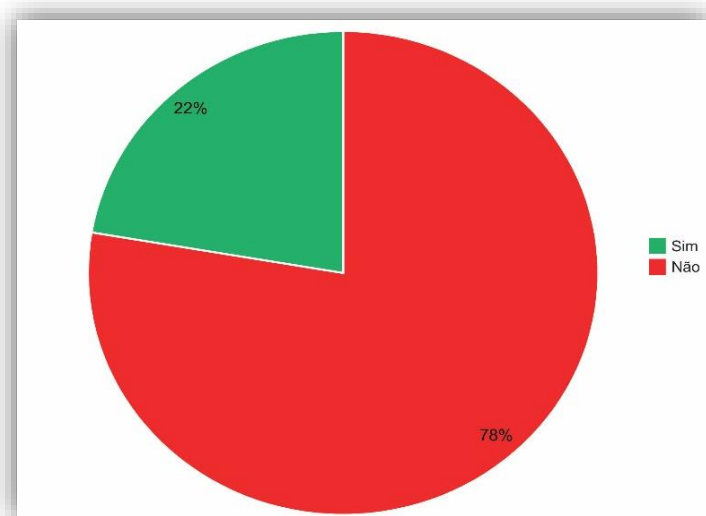


Figura 5 Sistema de qualidade no laboratório.

No Gráfico 5 (Fig. 6) pode-se verificar que, dos 27 respondentes que afirmaram possuir no laboratório o qual pertence um sistema da qualidade implementado, 9 disseram que tem BPL; 1 tem BPL e ISO 17025/2017; 1 tem BPL, ISO 17025/2017 e ISO 9000; 4 tem ISO 17025/2017; 4 tem ISO 9000; 2 tem ISO 9001, 1 tem CQB e 5 não souberam informar qual era o sistema implementado.

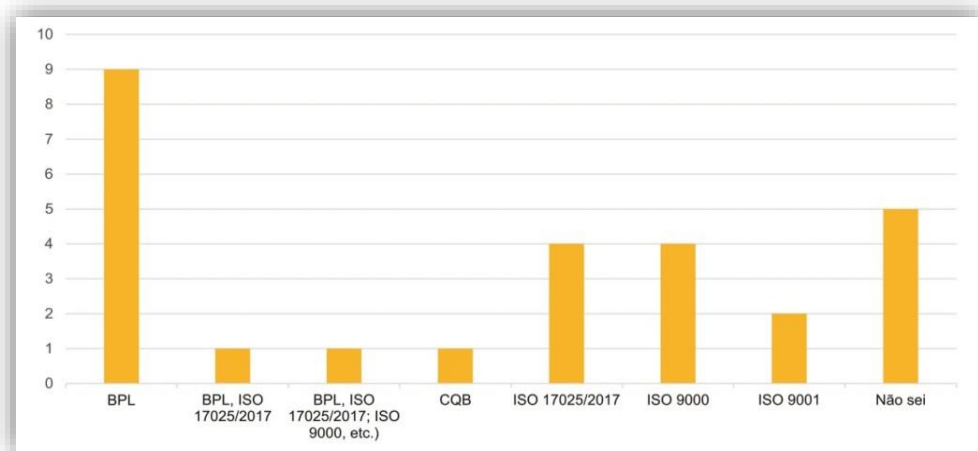


Figura 6 Sistema da qualidade implementado.

Quanto à existência ou não de disciplina relacionada a sistema de qualidade laboratorial durante a formação dos respondentes, verifica-se no Gráfico 6 (Fig. 7) que 77 (64%) responderam não existir disciplinas relacionadas ao sistema da qualidade durante sua formação e 44 (36%) responderam existir.

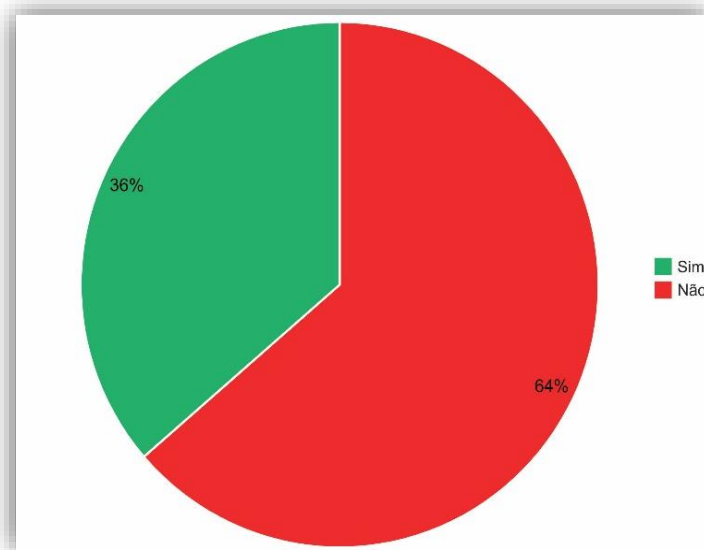


Figura 7 Disciplina sobre sistema da qualidade laboratorial.

No que se refere à faixa etária dos respondentes, Gráfico 7 (Fig. 8), tem-se que 74 (61%) tem entre 18 a 29 anos, 20 (17%) tem entre 30 a 39 anos, 10 (8%) tem entre 40 e 49 anos, 10 (8%) tem entre 50 e 59 anos, 5 (4%) tem entre 60 e 69 anos e 2 (2%) mais de 70 anos.

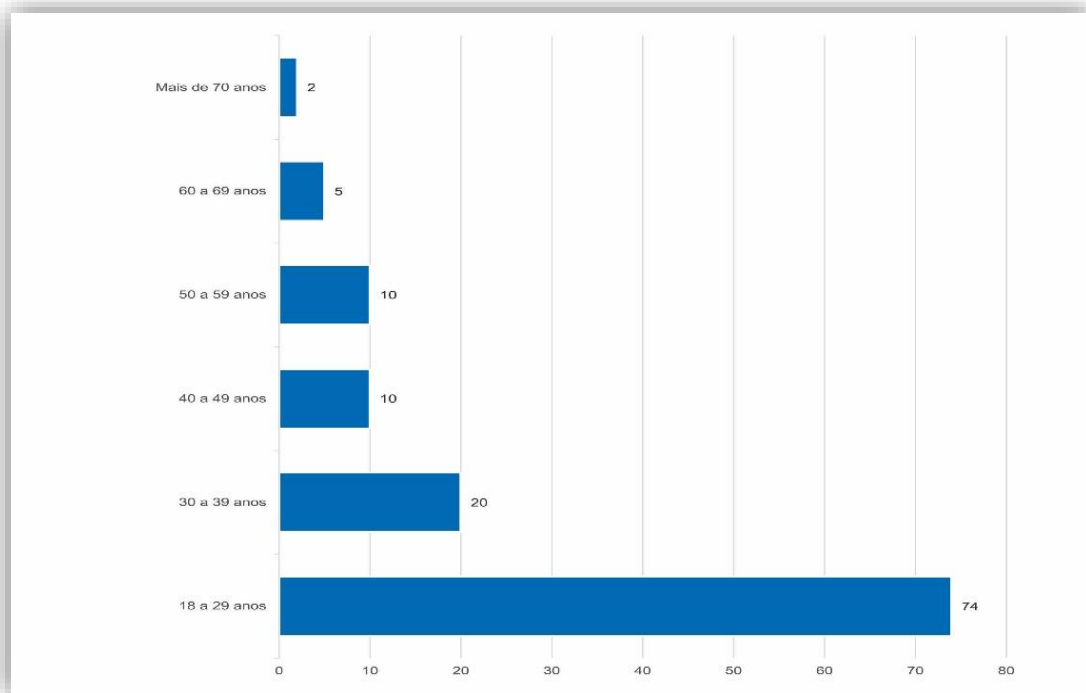


Figura 8 Faixa etária.

Quanto à titulação dos respondentes, eles declararam que, de acordo como Gráfico 8 (Fig. 9), 58 (48%) tem o grau de mestre, 44 (36%) tem graduação, 12 (10%) tem o grau de doutor e 7(6%) tem curso técnico de nível médio.

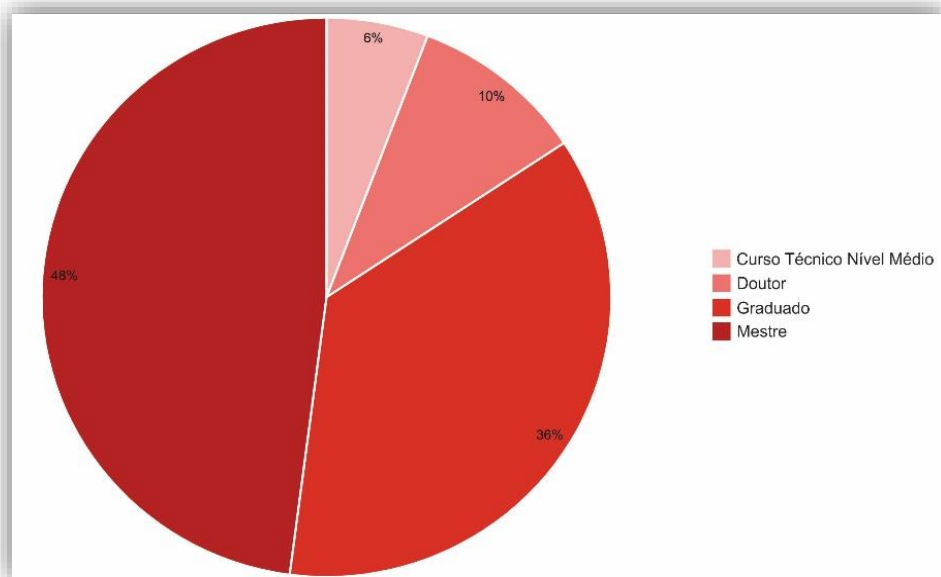


Figura 9 Titulação dos respondentes.

Sobre qual era o papel de cada respondente nos respectivos laboratórios, fica demonstrado no Gráfico 9 (Fig. 10) que: 54 (45%) eram doutorando, 35 (29%) eram mestrando, 12 (10%) iniciação científica, 6 (5%) Pesquisador Pós-doutor, 4 (3%) cientista responsável pelo laboratório, 4 (3%) gerente do laboratório, 3 (2%) técnico, 2 (2%) responsável da área técnica, 1 (1%) analista.

Discussão

O fato de 64% dos cento e vinte e um (121) participantes da pesquisa que fazem parte de cursos de pós-graduação em biotecnologia de várias instituições brasileiras, responderem não existir no seu curso (Biotecnologia) qualquer tipo de disciplina relacionada a Sistema da Qualidade Laboratorial durante a formação, evidencia a necessidade de disseminação do tema “Qualidade Laboratorial” e “Princípios das BPL”, por meio de disciplinas ou seminários nos cursos de pós-graduação em biotecnologia e outros cursos afins distribuídos no país. Esse tipo de capacitação é fundamental para esse segmento uma vez que ao se tornarem profissionais atuarão para garantir confiabilidade, qualidade e integridade nas atividades e processos laboratoriais, uma vez que a biotecnologia está totalmente relacionada ao uso de técnicas complexas envolvendo DNA / RNA, proteínas e

outras moléculas, cultura e engenharia de células e tecidos, vetores de genes e RNA, bioinformática, nanobiotecnologia [8].

Outrossim, reforçar a disseminação do tema BPL junto às redes de biotecnologia é estratégico, a julgar pela sua constituição que engloba instituições de ensino e pesquisa, bem como pelo foco das redes na formação de profissionais por meio da estruturação e consolidação de programas de pós-graduação em biotecnologia [2-6].

No que se refere à afirmação de que a grande maioria dos respondentes, que corresponde a 78% da amostra, não possui qualquer tipo de sistema de qualidade implementados nos laboratórios em que atuam e que estão ligados aos cursos de pós-graduação em biotecnologia, torna-se fundamental difundir o tema BPL para o público da biotecnologia, uma vez que a própria política de desenvolvimento da biotecnologia no país estabelece, através de Decreto, como ação estruturante a indução e o fortalecimento das instituições consideradas centros de excelência em testes clínicos, pré-clínicos e bioensaios, em conformidade com as normas internacionais de certificação de qualidade, com vistas a estimular o desenvolvimento regional nas áreas de apoio da bioindústria [1].

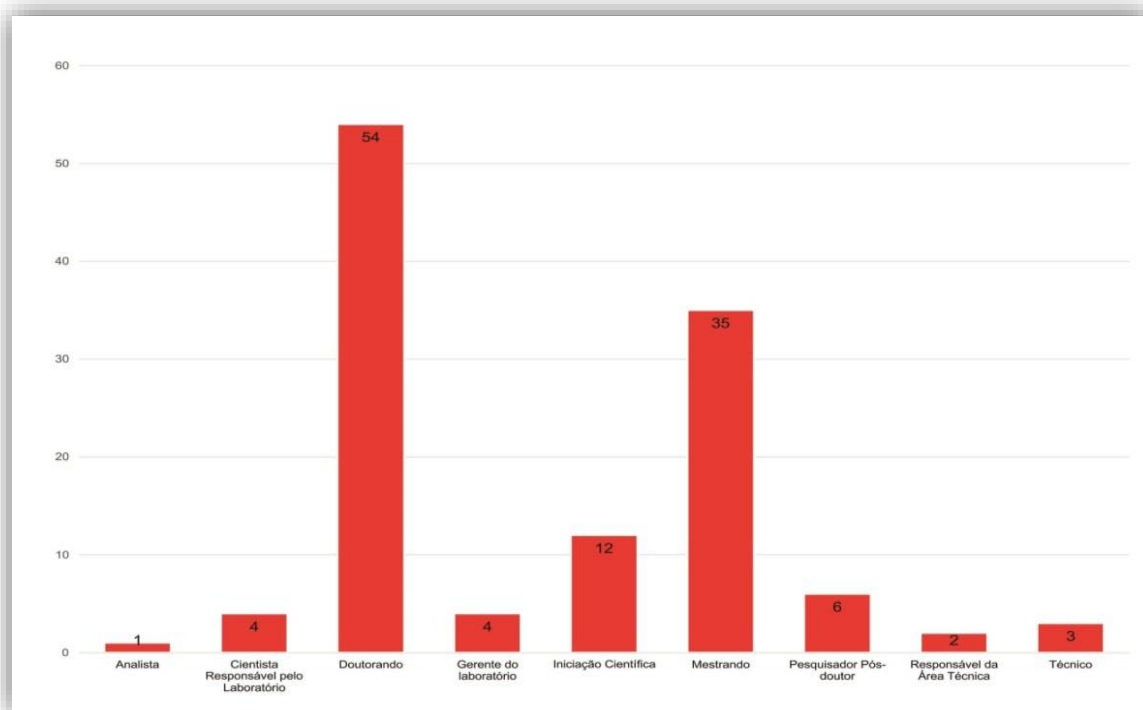


Figura 10 Papel dos respondentes no laboratório.

Importante ressaltar que mesmo para aqueles vinte e sete respondentes, os quais correspondem a 22% da amostra, que afirmaram possuir no laboratório um sistema da qualidade implementado, quando perguntado “Qual o sistema de qualidade implementado em seu laboratório?” demonstraram confusão ou desinformação do que é um sistema da qualidade laboratorial. Um respondente confundiu Certificado de Qualidade em Biossegurança (CQB) com BPL ou sistema de gestão de qualidade laboratorial. Cinco respondentes informaram não ter conhecimento se há ou não sistema de gestão da qualidade implementado no próprio laboratório em que atua.

Percebe-se, assim, que a capacitação e treinamento em BPL e sistema de gestão da qualidade é fundamental não somente para estudantes e profissionais que tem total desconhecimento do tema, mas também para aqueles que tem certo conhecimento mais que ainda necessitam se aprofundar no assunto para melhor sedimentar a compreensão já existente. Isso fica ainda mais evidente se considerarmos que, dentre os respondentes da pesquisa, quatorze (14) deles declararam ter funções diretamente ligadas ao laboratório, ou seja, quatro eram cientistas responsáveis pelo laboratório, quatro eram gerentes de laboratório, três eram técnicos no laboratório, dois eram responsáveis pela área técnica no laboratório e um era analista.

Associar a capacitação de profissionais do segmento da biotecnologia ao tema “qualidade laboratorial” é inevitável. Estudos recentes evidenciam o impacto financeiro da falta de reprodutibilidade na ciência, em particular, na biotecnologia. Estudos demonstraram que entre 50% a 90% de artigos publicados em revistas de impacto na área de biotecnologia não são reprodutíveis, provocando prejuízos de bilhões de dólares aos contribuintes e financiadores privados de pesquisa. Paralelamente ao custo, temos o impacto da falta de reprodutibilidade na credibilidade da ciência como um todo. As retratações e falhas na confirmação de dados corroem a confiança na capacidade da ciência efetivamente resolver os problemas da sociedade [22,23].

Para implementar um Sistema de Gestão da Qualidade é necessário haver ações de planejamento, preparação de documentação e treinamento de

recursos humanos. É considerado um processo lento e caro, mas pode trazer inúmeras vantagens à instituição como melhor imagem da empresa frente a seus clientes, padronização e maior controle dos processos, maior poder de detecção e correção de erros e maior confiabilidade analítica. No que se refere à implementação dos princípios da BPL aspectos críticos devem ser considerados como procedimentos e diretrizes que necessitam de interpretação durante a aplicação e a busca por melhoria contínua por meio do avanço do conhecimento técnico e científico, a fim de manter o sistema da qualidade. Quanto aos desafios é fundamental fazer com que pesquisadores e futuros pesquisadores acreditem nos benefícios desse processo por meio de treinamentos, pois estabelecer um Sistema de Gestão da Qualidade em pesquisa e desenvolvimento ou em qualquer outro segmento só é possível se as pessoas envolvidas estiverem convencidas ou acreditarem plenamente em suas vantagens [24].

Quando se opta em implementar normas de qualidade laboratorial, como a norma ISO 17025, em laboratórios de instituições de ensino, por exemplo, a escolha dos profissionais para funções relevantes deve ser realizada levando-se em conta a competência e habilidade técnica destes profissionais e seu comprometimento na busca pela qualidade em todos os aspectos. A escolha de alguém qualificado para executar e analisar criticamente todo o processo de medição, cálculo de incertezas e análise dos resultados finais, é fundamental para garantir resultados confiáveis em um laboratório. É comum nos laboratórios universitários determinadas funções serem exercidas por professores do quadro permanente, confiando estabilidade e continuidade às atividades e políticas do laboratório, porém, nada impede que alunos, bolsistas de iniciação científica, pesquisadores e estagiários possam atuar também em funções de apoio a este pessoal-chave. O envolvimento de professores e alunos de forma colaborativa nas atividades de gestão do laboratório é muito enriquecedora para ambos, mas principalmente para a formação profissional dos alunos e para a melhoria das condições do laboratório, além de ser fonte de muitos trabalhos técnicos e publicações científicas [25,26].

Conclusão

A pesquisa confirma o médio e baixo grau de percepção sobre o sistema da qualidade Princípios das BPL pelos estudantes de cursos de biotecnologia do país, evidenciando a necessidade da disseminação do tema Sistema da Qualidade Laboratorial não só para capacitar estudantes e profissionais relacionados ao segmento da biotecnologia, mas para todos aqueles com interesse em aprofundar o conhecimento sobre esse tema.

Associar a capacitação de estudantes e profissionais do segmento da biotecnologia ao tema “qualidade laboratorial” pode gerar ganhos evitando a falta de reprodutibilidade na ciência regulatória como um todo e, em particular, nos produtos e processos relacionados à biotecnologia.

Referências

- [1] BRASIL. Decreto no 6.041, de 8 de fevereiro de 2007. Institui a Política de Desenvolvimento da Biotecnologia, cria o Comitê Nacional de Biotecnologia e dá outras providências. **Diário Oficial da União**, 9 fev 2007.
- [2] REDE NORDESTE DE BIOTECNOLOGIA. **A Rede Nordeste de Biotecnologia**. Disponível em <http://www.renorbio.org/renorbio/sobre/a-rede-nordeste-de-biotecnologia>. Acesso em 17 mai. 2018.
- [3] BIONORTE. **Histórico**. Disponível em <https://bionorte.org.br/bionorte/rede-bionorte.html>. Acesso em 17 mai 2018.
- [4] UNIVERSIDADE FEDERAL DE GOIÁS. **Caracterização do ambiente da rede**. Disponível em <https://pgbb.prpg.ufg.br/n/67066-historico>. Acesso em 17 mai 2018.
- [5] REDE DE BIOTECNOLOGIA DA REGIÃO SUL. **Formação da Rede**. Disponível em <http://www.sulbiotec.com.br/pt/>. Acesso em 17 de mai 2018.
- [6] REDE NACIONAL DE PESQUISA EM BIOTECNOLOGIA MARINHA. **Projeto**. Disponível em <http://biotecmar.com.br/>. Acesso em 17 mai 2018.
- [7] GOMES, Ricardo Cruz; VICENTIN, Flávia Oliveira do Prado; GALINA, Simone Vasconcelos Ribeiro. **Redes de inovação das empresas de biotecnologia: mapeamento comparativo entre Brasil e Espanha**. In: XVI Congresso latino-iberoamericano de gestão de tecnologia., 2015, Porto Alegre, Anais, Porto Alegre: UFRGS, 2015, P.1-15.
- [8] DE MEIRA GUSMÃO, Alexandre Oliveira; DA SILVA, Antonio Rodrigues; MEDEIROS, Mauro Osvaldo. A biotecnologia e os avanços da sociedade. **Biodiversidade**, Mato Grosso, v. 16, n. 1, p.135-154, 2017.
- [9] FREIRE, Carlos Eduardo Torres. **Biotecnologia no Brasil: uma atividade econômica baseada em empresa, academia e Estado**. 2014. 201f. Tese (Doutorado em Sociologia)- Faculdade de Sociologia, Universidade de São Paulo, São Paulo, 2014.
- [10] ALVES, Nathália Guimarães; VARGAS, Marco Antônio; DE PAIVA BRITTO, Jorge Nogueira. Interações universidade-empresa: um estudo exploratório sobre as empresas de biotecnologia em saúde. **Revista Econômica**, Niterói, v. 20, n. 1, 2018. DOI: <https://doi.org/10.22409/economica.20i1.p388>.
- [11] DE OLIVEIRA, Henrique Sulzbach; SPENGLER, Rafael Luís. Inovações na área de biotecnologia em saúde humana em países em desenvolvimento e sua importância econômica e social: Uma reflexão sobre o cenário atual e perspectivas futuras. **Revista Caderno Pedagógico**, Lajeado, v. 11, n. 1, p.31-61, 2014.
- [12] PEREIRA, Rodrigues da Silva. Implantação de um sistema de gestão da qualidade em um laboratório de toxicologia baseado na norma ABNT NBR ISO/IEC 17025:2005. In: XII CONGRESSO NACIONAL DE EXCELÊNCIA EM GESTÃO, 2016, Rio de Janeiro, **Anais**, Rio de Janeiro: UFF, 2016, P.1-13.
- [13] DIAS, Valter Soares; BARQUETTE, Fernanda Rocha da Silva; BELLO, Alexandre Ribeiro. Padronização da qualidade: alinhando melhorias contínuas nos laboratórios de análises clínicas. **Revista Brasileira de Análises Clínicas**, Rio de Janeiro, v. 49, n. 2, p. 164-9, 2017.

- [14] JENA, G. B.; CHAVAN, Sapaná. Implementation of Good Laboratory Practices (GLP) in basic scientific research: Translating the concept beyond regulatory compliance. **Regulatory Toxicology and Pharmacology**, v. 89, p. 20-25, 2017.
- [15] COLLINS, Francis S.; TABAK, Lawrence A. Policy: NIH plans to enhance reproducibility. **Nature**, v. 505, n. 7485, p. 612-613, 2014. DOI:10.1038/505612a.
- [16] BEGLEY, C. Glenn; ELLIS, Lee M. Raise standards for preclinical cancer research. **Nature**, v. 483, n. 7391, p. 531-533, 2012. DOI: <https://doi.org/10.1038/483531a>.
- [17] INSTITUTO NACIONAL DE METROLOGIA, QUALIDADE E TECNOLOGIA. **NIT-DICLA-035: Princípios das boas práticas de laboratório (BPL)**. Rio de Janeiro: INMETRO, 2019.
- [18] WORLD HEALTH ORGANIZATION. **Handbook: Good Laboratory Practice (GLP)**. Quality practices for regulated non-clinical research and development. 2. Ed. Geneva: WHS, 2009.
- [19] ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. **Number I - OECD Principles on Good Laboratory Practice**. Paris: OECD, 1998.
- [20] INSTITUTO NACIONAL DE METROLOGIA, QUALIDADE E TECNOLOGIA. **Reconhecimento da Conformidade aos Princípios das BPL**. Disponível em http://www.inmetro.gov.br/monitoramento_BPL/reconhecimento_BPL.asp. Acesso em 18 abril 2018.
- [21] VIJAYARAGHAVAN, R. *et al.* **GLP (Good Laboratory Practice) Guidelines in Academic and Clinical Research: Ensuring Protection and Safety**. J Pharm Res Clin Pr, v. 4, p. 89-104, 2014.
- [22] BAKER, Monya. How quality control could save your science. **Nature News**, v. 529, n. 7587, p. 456, 2016. DOI: 10.1038/529456a
- [23] FREEDMAN, Leonard P.; COCKBURN, Iain M.; SIMCOE, Timothy S. The economics of reproducibility in preclinical research. **PLoS Biol**, São Francisco, 2015, 13.6: e1002165.
- [24] FRAGA, Hilda Carolina de Jesus Rios, et al. Avaliação da implementação de um sistema de qualidade em um laboratório de pesquisa básica: viabilidade e impactos. **Einstein**, São Paulo, v.10, n.4, p.491-497, 2012.
- [25] DE FELIPPES, Bruna Adriano; AGUIAR, Joselma Guimarães; DINIZ, Alberto Carlos Guimarães Castro. Sistema da qualidade em laboratórios universitários: incentivo ao ensino, pesquisa e extensão. **Revista de Ensino de Engenharia, Brasília**, v.30, n.2, p.14-23, 2012.
- [26] LAMERS, Marcelo Lazzaron; BARBOSA, Nathalia Mocellin; FOSSATI, Anna Christina Medeiros. Biotecnologia aplicada à Odontologia. IN: IX SALÃO DE ENSINO, 2013, Porto Alegre, **Anais**, Porto Alegre: UFRGS, 2013, p.1.

Inteligência Artificial e Metrologia



Charles B Prado

Centro de Capacitação (Cicma)

E-mail: cbprado@inmetro.gov.br

Metrologia é a ciência das medições. De acordo com VIM 2012 [1]:

"A metrologia engloba todos os aspectos teóricos e práticos da medição, qualquer que seja a incerteza de medição e o campo de aplicação."

Medições estão presentes em quase todos os aspectos de nossa sociedade, estão presentes em atividades como: controle de produção industrial, saúde, segurança de processos e produtos, qualidade de materiais, proteção ao consumidor e etc. Em torno de 4% a 6% do Produto Interno Bruto (PIB) de países desenvolvidos são investidos em metrologia [2].

É fundamental ter-se um **sistema de metrologia** que seja capaz de prover confiabilidade às medições. A confiabilidade é obtida através de uma cadeia de infraestrutura de qualidade capaz de obter dados e informações válidas em medições de alta precisão e relacioná-las a uma referência, documentando cada contribuição de incerteza de medição desta cadeia. De forma simples e resumida, é um processo de transferência de valores de medição de uma referência para instrumentos de medição comuns seguindo protocolos internacionais.

Estes protocolos internacionais, são procedimentos e métodos aprovados pela OIML (*International Organization of Legal Metrology*) que permitem que as transações comerciais (internamente ou externamente aos países) sejam executadas de forma justa, transparente e confiável.

No Brasil, o INMETRO (Instituto Nacional de Metrologia, Qualidade e Tecnologia) é o órgão responsável por executar as políticas nacionais de metrologia e da qualidade.

Dentro do contexto atual, aonde o mundo tem passado por transformações digitais constantes e em curtos espaços de tempo [3], a Metrologia tem um desafio de percorrer essa evolução e até propor inovações. Quando falamos de transformações digitais que afetam à sociedade atual nos referimos aos saltos tecnológicos advindos de tecnologias como sensores inteligentes e distribuídos, alta velocidade de comunicação de dados (5G, por exemplo), armazenamento e processamento de dados na nuvem, volume imenso de dados (*Big Data*), redes sociais, inteligência artificial, etc, só para citar alguns.

Imagine, hipoteticamente, um cenário em que dados de sensores distribuídos em uma linha de produção de uma indústria serão enviados e processados por um servidor local ou externo (na nuvem, por exemplo) e que o resultado final é um valor de medição associado aquele processo. Como garantir a confiabilidade e a rastreabilidade dessa medição ?

Enfim, os desafios são "desafiadores" permita-me a redundância. Agora, imagina quando nos deparamos com um dos temas mais usados na internet (*trend topics*), a IA "Inteligência Artificial" e nos perguntamos como e se ela irá influenciar todos esse universo da Metrologia. Será o nosso foco adiante, relacionar IA e Metrologia.

Primeiramente, temos que definir o que é Inteligência Artificial. Se procurarmos no Google encontraremos várias definições e muitas discussões do que é IA. Questionamentos sobre “ O que é Inteligência” e “qual a natureza do artificial” mostram que não é tão fácil se definir o que é IA. Usaremos a definição de Andreas Kaplan e Michael Haenlein [4], que é:

“Uma capacidade do sistema para interpretar corretamente dados externos, aprender a partir desses dados e utilizar essas aprendizagens para atingir objetivos e tarefas específicos através de adaptação flexível”

Mas, não podemos deixar de mencionar que a área de IA começou o seu desenvolvimento imediatamente depois da 2ª Guerra Mundial através do artigo “*Computing Machinery and Intelligence*” do famoso matemático inglês Alan Turing [5].

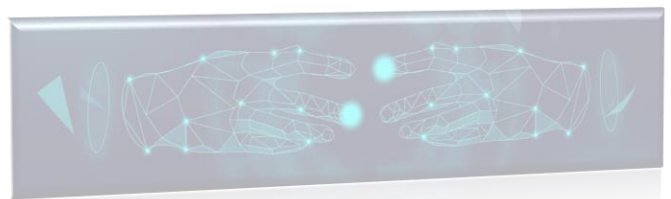
Hoje, com o avanço tecnológico, principalmente computacional, a área de IA se tornou viável e se expandiu rapidamente por toda a sociedade. É difícil não encontrar um setor da sociedade que não se utilize de ferramentas de IA para melhorar algum tipo de processo. Aplicações de IA são encontradas na indústria, por exemplo, para automação de processos industriais, na saúde no combate à pandemia da Covid-19, na economia quando se utiliza ferramentas de IA para fazer previsões de preços futuros na bolsa de valores, nos transportes, no setor aeroespacial, em jogos eletrônicos, no reconhecimento de padrões (face) em dispositivos como celulares e uma outra infinidade de aplicações nas mais diversas áreas. E, também na Metrologia.

No campo da Metrologia, o primeiro impacto que a IA vem trazendo é a melhoria dos processos internos. Neste sentido IA é vista como uma ferramenta capaz de melhorar a análise de dados, extrair características relevantes, otimizar processos. Por exemplo, IA pode ser usada para melhorar o processo de fiscalização de um determinado produto no mercado, através de um banco de informações desse produto - usar IA para detectar padrões que podem caracterizar, por exemplo, uma possível ação de fraude. IA poderá, também, ser usada para avaliar o “tempo de vida” de instrumentos de medição, como de medidores de energia elétrica, usando informações destes instrumentos como: potência, tensão, corrente e etc. Podemos resumir que IA para o INMs (Institutos Nacionais de Metrologia) tem uma aplicação direta no desenvolvimento de soluções para os seus problemas.

No entanto, alguns destes INMs são, também, institutos de tecnologia que realizam pesquisas avançadas que muitas vezes são bases de documentos de referência para diversos setores da sociedade. IA como base científica tem alguns desafios de estudo e pesquisa:

- ❑ **Acurácia** - IA como ferramenta de medição deve, necessariamente, ter um protocolo confiável e resultados avaliados para garantir a sua acurácia ;
- ❑ **Segurança** - IA é utilizado em áreas de tomada de decisão. Áreas como defesa cibernética, saúde, plantas industriais de geração de energia elétrica, são exemplos de áreas sensíveis cuja a utilização de IA deve vir seguida de uma avaliação de sua segurança à ataques cibernéticos;
- ❑ **Confiabilidade** - A confiabilidade da IA associada à capacidade de um determinado modelo operar constantemente como definido. Por exemplo, espera-se que uma determinada aplicação de IA que foi desenvolvida usando um conjunto de dados para treinamento, tenha o mesmo comportamento quando dados reais são apresentados.
- ❑ **Explicabilidade** - em muitas aplicações, IA é vista como caixa-preta, isto é, não se tem o conhecimento do que foi aprendido pela IA. No entanto, para algumas áreas, como de tomada de decisão em áreas estratégicas (geração de energia (*Smart Grid*), por exemplo) , é fundamental saber que informação ou conhecimento foi aprendido pela IA. É necessário “explicar” para o usuário o que a IA está usando de um conjunto de dados para tomar uma determinada decisão.
- ❑ **Robustez** - a robustez de uma IA é verificada através de sua resiliência à ataques que possam comprometer a sua segurança ou em induzir respostas erradas.

Cada um dos itens acima são linhas de pesquisa com enormes desafios, mas necessários para subsidiar o desenvolvimento da IA como base tecnológica para o avanço da Indústria 4.0 e da Internet das Coisas (*IoT*), só para citar duas áreas de grande relevância e de desenvolvimento atualmente.



Os desafios realmente são “desafiadores”. Os INMs tem um caminho claro à se perseguir quando se refere aos impactos da IA em seus negócios. Seja usando IA como ferramenta para melhorias de seus processos internos, seja como uma nova área de pesquisa. Podemos, de forma razoável, supor que mudanças estruturais e de formação de pessoal serão necessárias. Quanto mais cedo os INMs investirem no desenvolvimento e no uso de IA, mais rapidamente poderão atender às diversas demandas da sociedade que, em breve ou já agora, fazem uso de IA em seus negócios.

Referências

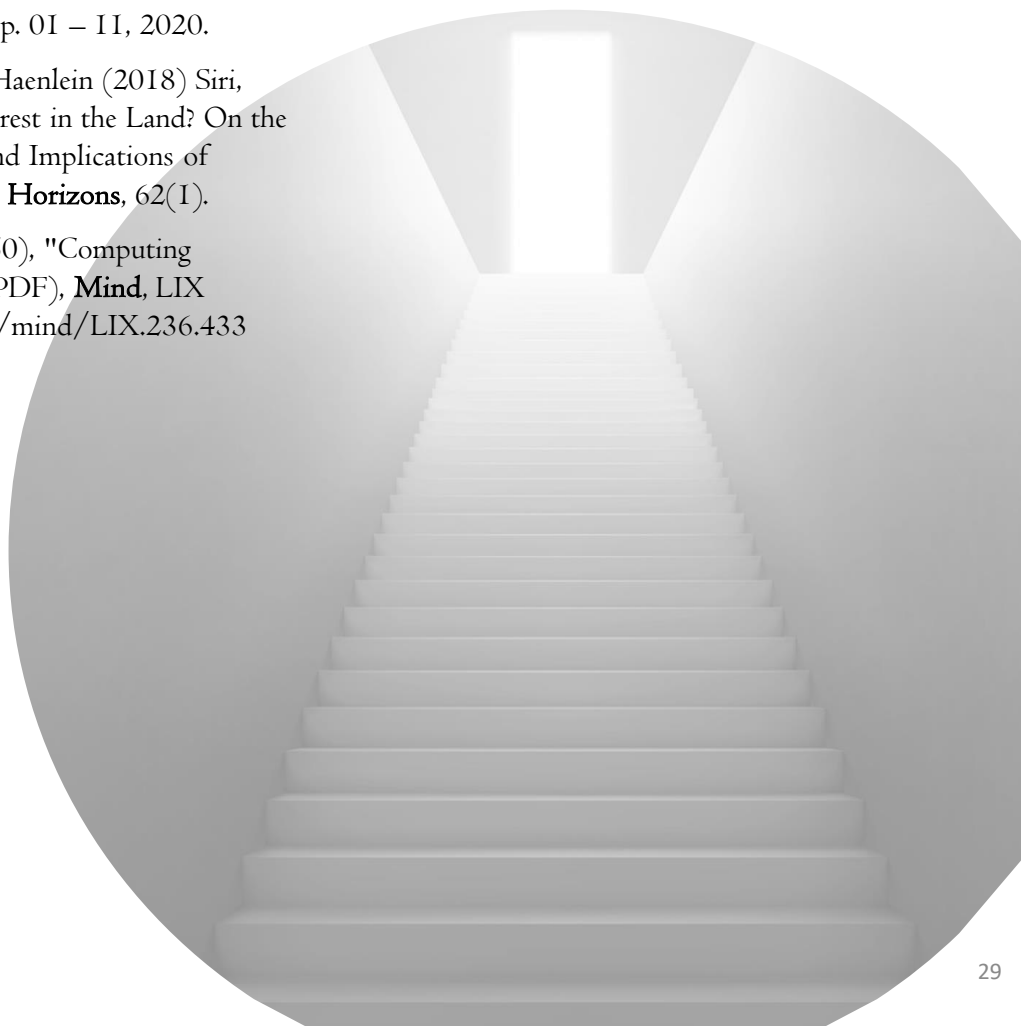
[1] Vocabulário Internacional de Metrologia – Conceitos fundamentais e gerais e termos associados (VIM 2012). Portaria Inmetro n.º 232, de 08 de maio de 2012.

[2] CBM, “Diretrizes estratégicas para a metrologia brasileira 2018-2022,” Tech. Rep., 2017. [Online]. Available:
<http://www.inmetro.gov.br/metcientifica/dirEstrategica/diretrizesEstrategicas.pdf>

[3] “Transformação digital no governo brasileiro: desafios, ações e perspectivas,” **Navus - Revista de Gestão e Tecnologia**, vol. 10, pp. 01 – 11, 2020.

[4] Andreas Kaplan; Michael Haenlein (2018) Siri, Siri in my Hand, who's the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence, **Business Horizons**, 62(1).

[5] Turing, Alan (October 1950), "Computing Machinery and Intelligence" (PDF), **Mind**, LIX (236): 433–460, doi:10.1093/mind/LIX.236.433





Big Data, Artificial Intelligence e Machine Learning no contexto da Análise de Dados em Data Science

Hércules de Melo Barcelos
Diretoria de Administração e Finanças (Diraf)
E-mail: hbarcelos@Inmetro.gov.br

Werickson F C Rocha
Diretoria de Metrologia Científica e Tecnologia (Dimci)
E-mail: wfrocha@inmetro.gov.br

1. Introdução

Existem muitos esforços despendidos ao longo dos anos na obtenção de metodologias mais abrangentes e que ajudem na interpretação de dados, os quais tornam-se numerosos em diversas áreas com o avanço da ciência e tecnologia, fazendo com que o conceito de *Data Science* seja bem mais robusto na atualidade. Neste contexto, o processo de análise de dados em *Data Science* une o domínio do conhecimento de um conjunto massivo de dados - que em termos gerais são denominados de *Big Data*, com formas modernas de tratamento de dados quem vão desde a utilização de conceitos simples de lógica de programação até a utilização de métodos complexos em Matemática, Estatística e de recursos oriundos da Ciência da Computação como a *Artificial Intelligence* (AI) ou Inteligência Artificial e o *Machine Learning* (ML) ou Aprendizado de Máquina que replicam a inteligência humana.

Em termos práticos, a *Data Science* busca transformar grandes grupos de dados em informações úteis através da construção de modelos capazes de realizar decisões complexas, para que isto seja possível, a *Data Science* possui como característica a multidisciplinaridade, apresentando-se como uma interseção entre diversas áreas do conhecimento. Através da Figura I apresentada por [1], pode-se visualizar as habilidades principais para um cientista de dados.



Figura I A Ciência de dados como área multidisciplinar.

É possível concluir através da Figura I que a interseção com diversas áreas permite ao cientista de dados construir modelos de análise de dados bem fundamentados e estruturados. O conhecimento em Estatística permite executar uma interpretação dos dados, possibilitando diferenciar a informação útil da informação não útil. A aplicação da Matemática fornece ferramentas capazes de auxiliar na construção de modelos mais complexos e a área do saber constitui a fonte de todo o dado a ser analisado e interpretado.

Com respeito à Ciência da Computação, esta disciplina colabora para a construção de diversas ferramentas computacionais tais como a AI e o ML, as quais permitem criar modelos eficazes durante a análise computacional de numerosos dados gerados.

As próximas seções deste trabalho apresentarão os conceitos básicos sobre *Big Data*, ML, AI e *Data Science*, possibilitando construir e aplicar um modelo preditivo de análise de dados fornecidos pela área médica para o diagnóstico do *Diabetes Mellitus*, utilizando como ferramenta de programação a linguagem R desenvolvida no ambiente do *software RStudio*.

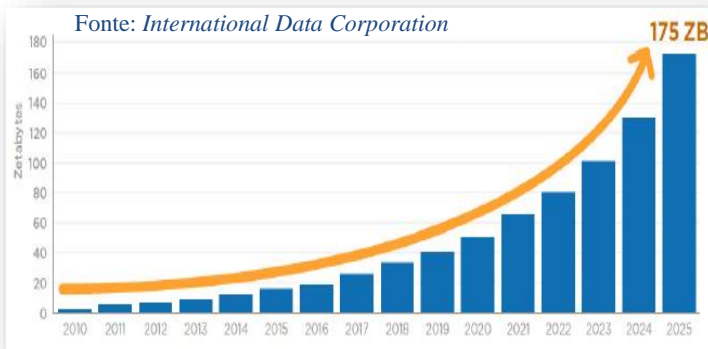


Figura 2 Crescimento de dados no mundo.

3. Artificial Intelligence

A partir de sua primeira edição, o livro “Inteligência artificial - uma abordagem moderna” se tornou um clássico na literatura sobre AI, neste livro Russel e Norvig [3] mencionam a seguinte frase “Inteligência artificial tem a ver com programas inteligentes, então vamos começar e escrever alguns”. De fato, em AI são desenvolvidos códigos computacionais que buscam simular um comportamento inteligente, tornando a AI uma ferramenta de grande utilidade em *Data Science*.

Dentro do domínio da AI coexistem vários campos de aplicação ainda mais profundos na geração e tratamento de dados, dentre eles têm-se o *Deep Learning* e o ML (Fig. 3). O *Deep Learning* [4] é um tipo específico de ML, formado por algoritmos de AI que podem realizar atividades impressionantes chegando a interagir como um ser humano tendo com o exemplo os *Chatbots* [5] que pela sua definição clássica, são programas de computador que processam uma linguagem natural de entrada de um usuário e geram respostas inteligentes que são então enviadas de volta para o usuário.

2. O conceito de Big Data

Associa-se o termo *Big Data* a um grande volume de dados que excede a capacidade de processamento de um banco de dados comum, chegando à dimensão de zetabytes de dados armazenados (Fig. 2).

Esta intensa geração de dados promove o desenvolvimento de tecnologias direcionadas ao armazenamento e processamento de grandes volumes de dados, tendo como exemplo o projeto *Apache Hadoop*, onde são produzidas bibliotecas de *software* em código aberto que permitem o processamento distribuído de grandes conjuntos de dados em *clusters* de computadores usando modelos de programação [2].

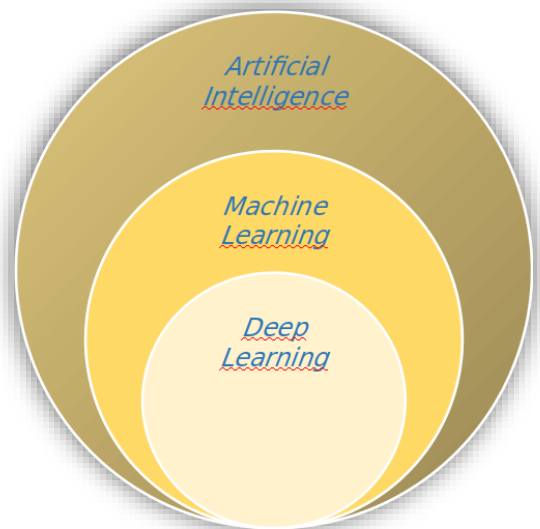


Figura 3 Diagrama de Venn par AI.

4. Definindo Machine Learning

Em 1952, o Engenheiro Arthur Samuel desenvolveu um protocolo em um jogo de damas que fez o jogo aprender com os erros e acertos do usuário, com isto, em 1959 Arthur Samuel definiu o aprendizado de máquina como um "campo de estudo que dá aos computadores a capacidade de aprender sem serem programados explicitamente" [6]. Em termos técnicos, entende-se que o ML é um subconjunto atrelado ao conceito de AI (Fig. 3), sendo direcionado para a criação de computadores que pensem como os humanos, permitindo a tomada de decisões através de algoritmos que reconhecem padrões de resposta. Estes algoritmos são construídos com base em três tipos de aprendizagem, a supervisionada, não supervisionada e por reforço.

4.1 Aprendizagem supervisionada

A construção de modelos através de Algoritmos de ML em com aprendizagem supervisionada tem como objetivo prever o valor da variável de saída y (considerada com o vetor resposta), com base em várias variáveis de entrada \mathbf{X} (armazenadas em forma matricial) [7]. Este tipo de aprendizagem será utilizado na construção dos modelos preditivos deste trabalho e alguns métodos estatísticos como o *K-Nearest Neighbors* (KNN), *Naive Bayes* e *Support Vector Machine* (SVM) são bons exemplos para a criação de modelos supervisionados.

4.2 Aprendizagem não supervisionada

O foco da aprendizagem não supervisionada é clusterizar os dados, tendo como base certas características similares, dispensando apresentar ao algoritmo dados rotulados de saída, ou em outras palavras, objetiva-se aqui descrever as associações e padrões entre um conjunto de variáveis de entrada sem uma variável de saída [7]. Este modelo é recomendado para problemas de agrupamento ou para a associação e detecção de anomalias. Um bom exercício seria a criação de um modelo que identifique anomalias ou agrupe clientes com base em comportamentos similares. Os algoritmos *K-means clustering*, Regras de Associação e Análise de Componentes Principais (PCA) são exemplos de ferramentas para a criação de máquinas de aprendizagem não supervisionada.

4.3 Aprendizagem por reforço

O algoritmo de ML desenvolvido em aprendizagem por esforço tem como princípio o aprendizado com base na interação com o ambiente, não sendo apresentadas previamente as ações que devem ser tomadas. Este tipo de código computacional consegue aprender comportamentos muito complexos sem precisar de nenhum dado de treinamento rotulado [8], possibilitando ao modelo criado por este algoritmo tomar decisões de curto prazo enquanto otimiza para um objetivo a longo prazo. Os métodos *Q-learning*, *Sarsa*, *Q(Lambda)* [9] fazem parte deste tipo de aprendizagem.

5. Construindo modelos preditivos em *Data Science*

Neste tópico foram desenvolvidos 3 modelos preditivos do tipo supervisionado para identificar os padrões possíveis para a realização de diagnósticos do *Diabetes Mellitus* [10]. O Volume de dados selecionados para o desenvolvimento deste modelo corresponde a um total de 712 amostras associadas a pacientes do sexo feminino, contendo resultados positivos e negativos para o vetor y e informações associadas à estes resultados para a matriz \mathbf{X} . Com relação à veracidade dos dados, estes fazem parte do famoso *Dataset* (conjunto de dados tabulados) sobre a incidência de Diabetes em mulheres pertencentes à etnia dos índios Pima [11] um pequeno grupo étnico do Sudoeste dos Estados Unidos e a fonte geradora desses dados está localizada em um grande acervo fornecido pelo *National Institute of Diabetes and Digestive and Kidney Diseases* [12]. A ideia principal é que o modelo apresente uma acurácia de no mínimo 75%, cabendo às disciplinas de Ciência da Computação, Matemática e Estatística o trabalho de dar forma aos Algoritmos de ML, base para a construção dos modelos deste artigo. O ambiente de desenvolvimento utilizado foi o *Rstudio*, um software livre e integrado para linguagem R utilizada neste trabalho.

5.1 Características dos dados

Os dados do *Dataset* são do tipo estruturados [13] e foram acessados na comunidade de ML e *Data Science* chamada *Kaggle* [14]. A Tabela 1 apresenta os dados das primeiras 5 amostras do *Dataset* e utilizando a forma de descrição dos dados em *Data Science* [15], as linhas da Matriz **X** serão identificadas como “instâncias”, as colunas da Matriz **X** serão chamadas de Variáveis **X**'s. O vetor **y** será chamado de Variável **Y** e representa o diagnóstico das pessoas que desenvolveram ou não a doença, sendo 1 para diagnósticos positivos e 0 para negativos. A Tabela 2 identifica cada Variável **X** da Matriz **X**.

Tabela 1 *Dataset* para a construção do modelo de *Machine Learning*.

Amostras	Matriz X								Vetor y
	Gravidez	Glicose	Pressão Sanguínea	Espessura de pele	Insulina	IMC	F*	Idade	Resultado
1	6	148	72	35	0	33,6	0,627	50	1
2	1	85	66	29	0	26,6	0,351	31	0
3	8	183	64	0	0	23,3	0,672	32	1
4	1	89	66	23	94	28,1	0,167	21	0
5	0	137	40	35	168	43,1	2,288	33	1

* Função de Pedigree do Diabetes

Fonte: *Kaggle* (2020).

Tabela 2 Dados sobre as 8 Variáveis da Matriz **X**.

Variável X	Descrição	Unidade	Classificação do tipo de dado no <i>RStudio</i>
Gravidez	Número de vezes que engravidou	Não aplicado	Inteiro
Glicose	Concentração de glicose plasmática a duas horas em um teste oral de tolerância à glicose	mg/dL	Inteiro
Pressão Sanguínea	Pressão arterial diastólica	mmHg	Inteiro
Espessura de Pele	Espessura da dobra cutânea do tríceps	mm	Inteiro
Insulina	Insulina sérica de 2 horas	mUI/ml	Inteiro
IMC	Índice de massa corporal	kg/m ²	Não inteiro
Função de Pedigree do Diabetes	Função relacionada à fatores genéticos [18]	Não aplicado	Não inteiro
Idade	Idade do indivíduo	Anos	Inteiro

Fonte: *Kaggle* (2020).

5.2 Análise exploratória

Em *Data Science* é muito comum realizar inicialmente a fase de Análise exploratória dos dados, ela ajuda a identificar problemas comuns que ocorrem em um *Dataset*, como por exemplo a falta de componentes em cada Variável X ou Y, também pode ocorrer problemas com as componentes como algum erro de digitação numérica que descaracterize o dado. Essas análises foram realizadas e não foi encontrada nenhuma anomalia nos dados.

Continuando com a análise dos dados foi possível utilizar a Estatística descritiva [16] e graficar no *RStudio* a Figura 4, nela pode-se verificar que os diagramas de caixa apresentam poucos dados discrepantes com relação às Variáveis X. É importante destacar que a aplicação de qualquer condição de filtro dos dados irá interferir no resultado final do modelo, sendo necessário consultar especialistas na área médica antes de realizar qualquer análise discriminatória dos dados.

5.3 Construção do modelo

Como apresentado na Figura 5 [17], a criação do modelo preditivo passa por 4 fases, a fase inicial chamada de pré-processamento contempla toda a análise exploratória dos dados e a separação dos dados (passo I da Figura 5) em dois novos *Datasets* chamados de Dados Treino e Dados Teste. A técnica para criação do modelo exige o seu treinamento e teste para a sua validação final, para isto foi utilizado o pacote *caTools* e a função *sample.split* do *RStudio* onde foi possível definir o valor percentual de 70% totalizando 498 amostras para os Dados Treino e 214 para os Dados Teste.

Na fase 2 (passo 2 da Figura 5), o processo de aprendizagem de máquina solicita qual Algoritmo de ML deverá ser utilizado, como o vetor y é conhecido foi adotado o tipo de aprendizagem supervisionada e foram selecionados os Algoritmos de ML para problemas de classificação, pois o objetivo final é classificar a novas amostras como 0 ou 1. Para este passo foi utilizada a função “*train*” do pacote “*Carret*” no *RStudio*, inserindo nesta função a Variável Y, o agrupamento de todas as Variáveis X predictoras (vide Tabela I), o *Dataset* com os Dados Treinos e o Algoritmo de *Machine Learning* definido. Primeiramente foi escolhido o Algoritmo de ML *K-Nearest Neighbor* (KNN) [18].

Este método se baseia nos vizinhos mais próximos e ao ser apresentado às Variáveis X do Dado treino ele irá classificar o resultado para cada amostra e a sua acurácia será medida. O Método KNN utiliza o parâmetro k para identificar a quantidade de vizinhos mais próximos que deve ser considerada pelo algoritmo, desta forma o modelo KNN foi construído considerando um intervalo inteiro de 1 a 20 para k , gerando assim os resultados apresentados na Figura 6 (a).

Ao seguir para a fase de avaliação (Fig. 5), foi utilizado como métrica de desempenho a acurácia encontrada para cada k utilizado, e o melhor valor de acurácia encontrado foi de 72,35909% (Fig. 6 (a)) com coeficiente Kappa (parâmetro do Método) igual a 0,33257 para $k=19$, contudo não foi atendido a meta de 75% de acurácia na etapa de treinamento para este primeiro modelo.

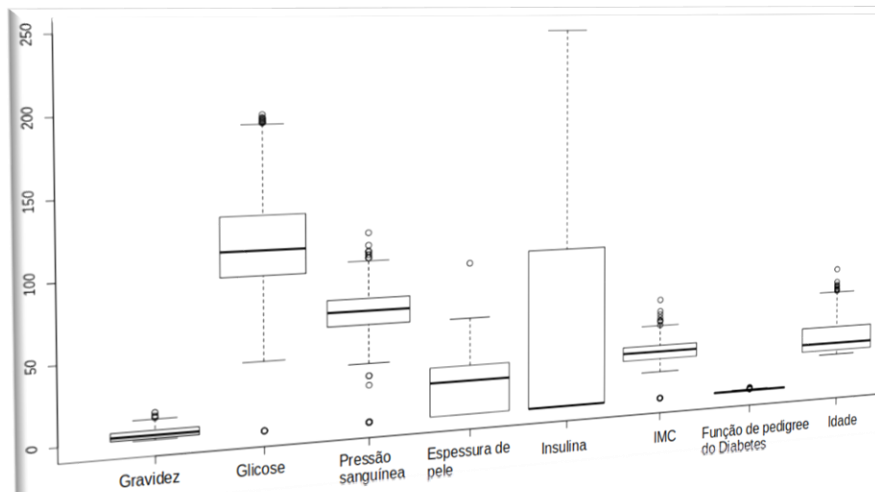


Figura 4 Diagrama de caixa para as Variáveis X do *Dataset*.

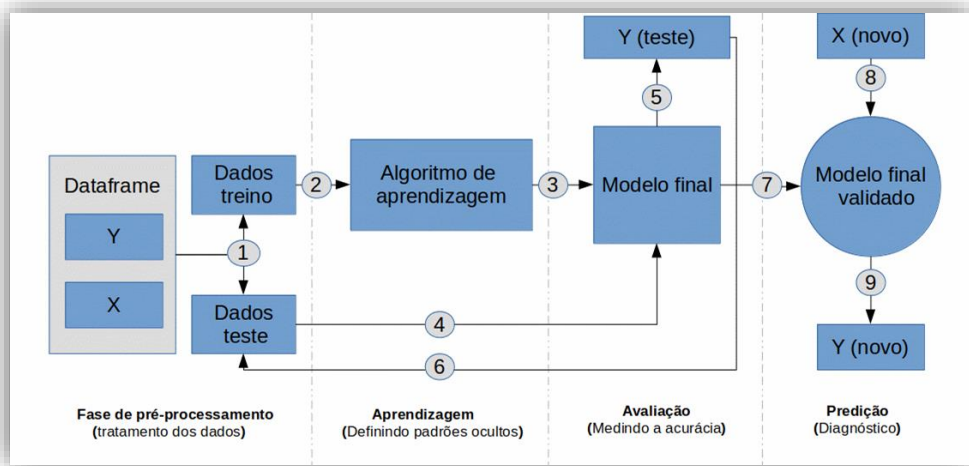


Figura 5 Modelo preditivo de *Machine Learning* com aprendizagem supervisionada.

As fases anteriores foram repetidas para o segundo modelo utilizando agora o Algoritmo ML *Naive Bayes* [19]. Este Algoritmo foi construído seguindo a técnica de classificação baseada no teorema de Bayes com uma suposição de independência entre os preditores [20]. Resumidamente, um Algoritmo do método de *Naive Bayes* assume que a presença de uma característica particular em uma classe não está relacionada com a presença de qualquer outro recurso. Por exemplo, um animal pode ser um cachorro se tiver calda, patas e ser domesticável. Mesmo que esses recursos dependam uns dos outros ou da existência de outras características, todas estas propriedades contribuem de forma independente para a probabilidade de que esse animal seja um cachorro, é por isso que o método é conhecido como '*Naive*' (ingênuo). Repetindo o que foi feito para o método utilizado anteriormente, foi encontrado para este segundo modelo o valor de 75,80241% para a acurácia e coeficiente Kappa de 0,44730 (Fig. 6(b)) sendo atingida a meta de 75% de acurácia para este exemplo.

Prosseguindo para a criação do terceiro modelo, foi aplicado agora o Algoritmo ML chamado de *Support Vector Machine* (SVM) [21]. Em síntese o SVM é utilizado para realizar regressões e classificações gerais (do tipo nu e épsilon), bem como estimativas de densidade. O desenvolvimento deste método faz a utilização de *Radial Base Functions* (RBF) [22], muito utilizada em processos de interpolação aplicados em Métodos numéricos [23,24]. Pela Figura 7 foi possível encontrar para o terceiro modelo uma acurácia de 76,48089% para um valor de Sigma de 0,03495, variável Cost (Constante do termo de regularização da formulação de Lagrange) igual a 0,50 e coeficiente Kappa igual a 0,43912.

Após o treinamento dos 3 modelos, foi constatado que os modelos baseados em *Naive Bayes* e SVM atenderam a meta estabelecida de 75% de acurácia na etapa de treinamento (Tabela 3), contudo os 3 modelos aprenderam como as variáveis se relacionam e desta forma eles seguiram para a aplicação do *Dataset* de Dados Teste, onde agora foram realizados os passos de 4 a 6 da Figura 5.

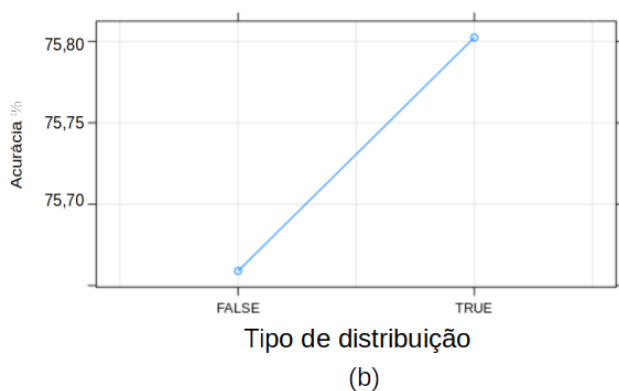
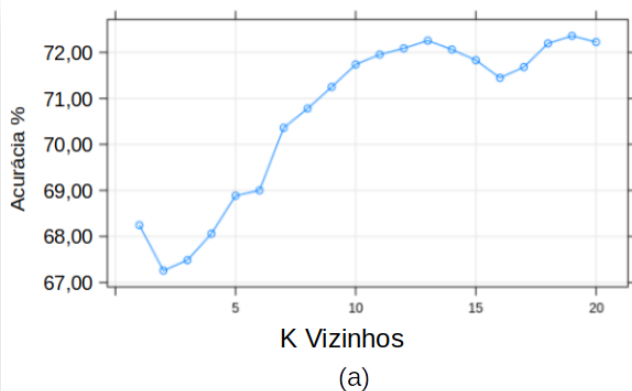


Figura 6 (a) Curva de acurácia com a aplicação do Algoritmo ML tipo KNN. (b) Método *Naive Bayes*.

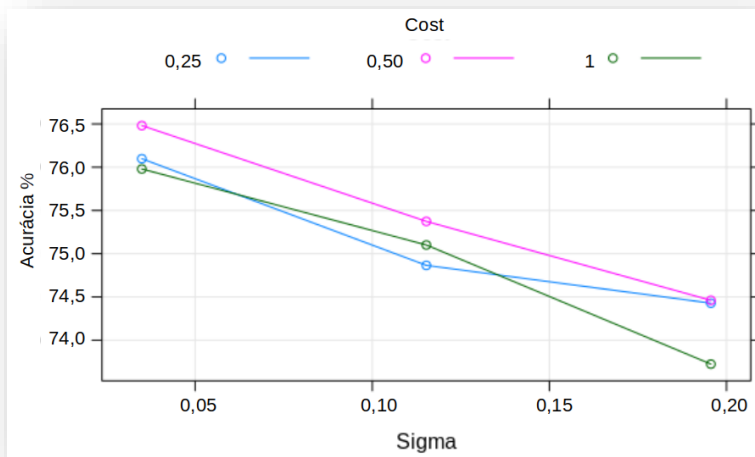


Figura 7 Método SVM.

Esse novo procedimento permitiu realizar previsões e medir a acurácia de cada modelo relacionada com o quanto foi aprendido sobre os dados treino. Para realizar as previsões, cada modelo recebeu como dados de entrada as Variáveis X dos Dados teste e com isto, cada modelo entregou como resposta os valores previstos para a Variável Y dos Dados teste. Os valores previstos foram comparados com os valores reais e a acurácia de cada modelo foi registrada na Tabela 4. Para realizar a tarefa de previsão no *RStudio* foi utilizado o comando “*Predict*” e acurácia foi obtida através do comando “*confusionMatrix*”.

Tabela 3 Acurácia encontrada para os métodos aplicados com dados treino.

Método		Acurácia (dados treino)	Coefficiente Kappa
<i>K-Nearest Neighbor</i>	KNN	72,35909%	0,33257
<i>Naive Bayes</i>	<i>Naive Bayes</i>	75,80241%	0,44730
<i>Support Vector Machines</i>	SVM	76,48089%	0,4391171

A Tabela 4 permite concluir que o todos os modelos apresentaram acurácia acima de 75% e que o Modelo construído com base no algoritmo SVM superou os demais, demonstrando um melhor aprendizado sobre os dados fornecidos.

Tabela 4 Acurácia encontrada para os métodos aplicados com dados teste.

Método		Acurácia (dados teste)	Coefficiente Kappa
<i>K-Nearest Neighbor</i>	KNN	76,17%	0,4179
<i>Naive Bayes</i>	<i>Naive Bayes</i>	75,70%	0,4171
<i>Support Vector Machines</i>	SVM	79,91%	0,5092

De posse dos 3 modelos previamente validados, foram criadas 5 amostras de pacientes fictícios conforme a Tabela 5 e submetidas aos 3 modelos, gerando assim as previsões dos diagnósticos para o desenvolvimento da *Diabetes Mellitus* conforme a Tabela 6.

Observando a Tabela 6 e considerando como referência o terceiro modelo SVM, verifica-se que o primeiro modelo KNN apresentou resultados mais próximos ao modelo de referência.

Tabela 5 Amostras criadas para a realização de simulações de diagnósticos positivos ou negativos para a doença de Diabetes.

Amostras	Variáveis X							
	Gravidez	Glicose	Pressão Sanguínea	Espessura de pele	Insulina	IMC	F*	Idade
1	0	171	77	23	94	27,5	1,627	23
2	3	97	55	12	167	24,3	2,351	37
3	6	210	61	40	146	37,7	0,372	44
4	9	67	45	15	101	29,1	0,267	47
5	12	101	83	45	85	49,1	2,08	54

* Função de Pedigree do Diabetes

Tabela 6 Simulações de diagnósticos para a doença de diabetes obtidos pelos 3 modelos construídos.

Amostras	Diagnóstico (Variável Y ou Resposta)		
	KNN	Naive Bayes	SVM
1	<i>Positivo</i>	<i>Negativo</i>	<i>Positivo</i>
2	<i>Negativo</i>	<i>Positivo</i>	<i>Negativo</i>
3	<i>Positivo</i>	<i>Positivo</i>	<i>Positivo</i>
4	<i>Negativo</i>	<i>Negativo</i>	<i>Negativo</i>
5	<i>Negativo</i>	<i>Positivo</i>	<i>Positivo</i>

6. Conclusões

A partir do exposto, ficou claro que o desenvolvimento de trabalhos em *Data Science* envolve uma abordagem multidisciplinar, requerendo daqueles que atuam nesta área o conhecimento e domínio de modernas técnicas e ferramentas tecnológicas para a análise e interpretação de dados, garantindo desta forma uma clara informação sobre a previsão e compreensão do volume de dados gerados em um processo complexo.

A simulação experimental foi importante para materializar um aplicação de análise de dados, sendo possível construir Modelos preditivos e aplicá-los em novas condições. Com respeito aos Algoritmos utilizados, o SVM superou os demais durante as fases de treino e teste dos modelos, contudo o Algoritmo KNN apresentou uma boa aproximação das previsões quando comparado com o método SVM, o que já era previsto com base na Tabela 4. Cabe mencionar que as instâncias apresentadas para a realização dos diagnósticos simulados foram criadas de forma aleatória, onde para casos reais as Variáveis X podem estar correlacionadas e isto não foi considerado.

Para realização de trabalhos futuros, torna-se interessante testar outros Métodos de *Machine Learning* disponíveis como por exemplo o *Decision Trees*, *Neural Networks*, etc., além de utilizar dados de amostras reais na etapa simulação dos diagnósticos.

7. Referências

- [1] Cuesta H, Kumar S 2016 Practical Data Analysis (ed. Birmingham: Packt Publishing).
- [2] Hadoop: O que é, conceito e definição 2016, Cetax [s. l.]. Disponível em: <https://www.cetax.com.br/blog/apache-hadoop/>. Acesso em: 14 nov 2020.
- [3] Russel S, Norvig P 2020 Artificial Intelligence: A Modern Approach. (4. ed. Pearson Series).
- [4] Goodfellow I, Bengio Y, Courville A 2017 Deep Learning (MIT Press).
- [5] Khan R, Das A 2018 Build Better Chatbots (Apress).
- [6] Reza M. Machine Learning 1995 Pittsburg PA: CMUCC. Disponível em: <http://www.contrib.andrew.cmu.edu/~mndarwis/M.L.html>. Acesso em: 25 out. 2020.
- [7] Rocha W F C, Prado C B, Blonder N 2020 Comparison of Chemometric Problems in **Food Analysis using Non-Linear Methods**. *Molecules*, 25: I-55.
- [8] Borini G 2018 Reinforcement Learning: Aprenda com o mini carro autônomo DeepRacer, cio [s. l.]. Disponível em: <https://cio.com.br/tendencias/reinforcement-learning-aprenda-com-o-mini-carro-autonomo-deepracer/>. Acesso em: 15 nov 2020.
- [9] Monteiro S T, Ribeiro C H C 2003 Desempenho de Algoritmos de Aprendizagem por Reforço sob condições de ambiguidade sensorial em robótica móvel. **Revista Controle e Automação**, 5: 320-338.
- [10] Oguntibeju O O 2013 Diabetes Mellitus – Insights and perspectives (InTech Press Team).
- [11] Seliar, M 2009 Diabetes: Aprendendo com os índios, Academia [s. l.]. Disponível em: <https://www.academia.org.br/artigos/diabetes-aprendendo-com-os-indios>. Acesso em: 10 nov. 2020.
- [12] NIH: National Institute of Diabetes and Digestive and Kidney Diseases 2020, Kniddk.nih, U. S. Department of Health and Human Services. Disponível em: <https://www.niddk.nih.gov/health-information/diabetes>. Acesso em: 10 nov. 2020.
- [13] A diferença entre dados estruturados e não estruturados, docmanagement [s. l.]. Disponível em: <https://docmanagement.com.br/03/06/2015/a-diferenca-entre-dados-estruturados-e-nao-estruturados/>. Acesso em 15 nov 2020.
- [14] Data: Diabetes Data Set 2020, Kaggle [s. l.]. Disponível em: <https://www.kaggle.com/mathchi/diabetes-data-set>. Acesso em: 10 nov. 2020.

- [15] Introdução à Análise de Dados 2020 Fundação Escola Nacional de Administração Pública. Brasília: ENAP. Disponível em: <https://www.escolavirtual.gov.br/catalogo#inicio-listagem-cursos>. Acesso em: 25 out. 2020.
- [16] Morettin P A, Bussab W O 2004 Estatística Básica (Ed. Saraiva).
- [17] Raschka S 2015 Python Machine Learning (Birmingham: Packt Publishing,).
- [18] Shakhnarovich G, Darrell T, Indyk P 2006 Nearest-Neighbor Methods in Learning and Vision (MIT Press).
- [19] Herrmann, D, Wendolsky R, Federrath H A. 2009 Attacking Popular Privacy Enhancing Technologies with the Multinomial Naïve-Bayes Classifier, **CCSW'95**, nov 2009 Disponível em: <https://dl.acm.org/doi/10.1145/1655008.1655013>. Acesso em 12 novembro 2020.
- [20] Rai, S. 2016 6 passos fáceis para aprender o algoritmo Naive Bayes (com o código em Python). VOODOO.PRO, 2016. Disponível em: <https://www.voodoo.pro/insights/6-passos-faceis-para-aprender-o-algoritmo-naive-bayes-com-o-codigo-em-python>. Acesso em: 13 nov. 2020.
- [21] Meyer D 2020 SVM: Support Vector Machines, Rdocumentation, 2020. Disponível em: <https://www.rdocumentation.org/packages/e1071/versions/1.7-4/topics/svm>. Acesso em: 13 nov. 2020.
- [22] Buhmann M D 2003 Radial Basis Function: Theory and Implementations (Cambridge University Press).
- [23] Barcelos H. M.; Loeffler, C. F. The direct interpolation boundary element method applied to smoothly inhomogeneous Laplace's problems. **Eng. Anal. Bound. Elem.**, V. 105, n. 1, p. 155-164, 2019.
- [24] Barbosa J P, Loeffler C F 2020 Application of Boundary Element Method superposition technique for solving natural frequencies in piecewise homogeneous domains. *Comput. Math. with Applications*, 79: 1131–1144.

Inteligência Artificial Aplicada a Metrologia - Uma ferramenta de Calibração

Elizeu Calegari
Diretoria de Metrologia Científica e Tecnologia (Dimci)
E-mail: ecalegari@inmetro.gov.br

Werickson F C Rocha
Diretoria de Metrologia Científica e Tecnologia (Dimci)
E-mail: wfrocha@inmetro.gov.br
Gabriel Sarmanho
Diretoria de Metrologia Científica e Tecnologia (Dimci)
E-mail: gsarmanho@inmetro.gov.br

1. Introdução

O foco da Diretoria de Metrologia Científica (DIMCI) do Inmetro é a pesquisa científica voltada para ensaios, calibrações e a produção de material de referência certificado (MRC). A rastreabilidade é outro aspecto fundamental e envolve uma cadeia ininterrupta e documentada de calibrações, cada uma contribuindo para a incerteza de medição (conforme o Vocabulário Internacional de Metrologia – VIM). Para desenvolver essas pesquisas, o pesquisador necessita de ferramentas de suporte para execução dos cálculos necessários e também a obtenção e apresentação dos resultados. Atualmente diversos programas com base estatística foram desenvolvidos e que tornaram mais eficientes o tratamento de dados. Existem diversas ferramentas de suporte, tais como o **Excel**, **SPSS**, o **BMDP**, **Minitab**, **NCSS**, **SYSTAT**, **Statplus**, **SAS**, entre diversos outros, além ainda de *softwares* estatísticos com código livre e gratuito, como o é o caso do **R**. O uso dessas ferramentas apresenta alguns desafios como a aquisição das licenças necessárias, customização para atender necessidades específicas, e inclusive o valor agregado importantíssimo de possuir um domínio intelectual no desenvolvimento dessas tecnologias. A DIMCI dispõe de um grupo de pesquisadores que vem desenvolvendo um esforço nesse sentido e a ferramenta apresentada neste trabalho é um exemplo.

A calibração de instrumentos de medição é um procedimento que permite identificar se

instrumentos utilizados na indústria estão aptos para o processo de medição. O objetivo é evitar que valores errôneos ou imprecisos apresentados por esses instrumentos venham interferir na qualidade final do produto. Em um sistema de medição (Fig. 1), pretende-se conhecer um modelo da relação de dependência da saída $y(t)$ em função da entrada $x(t)$. Existem dois métodos principais de determinação do modelo: o **analítico** – baseado em princípios da físicos, químicos ou biológicos; e o **empírico** (ou heurístico – que visa identificar o processo via experimental. A calibração se baseia no método empírico. Os modelos empíricos são menos complexos, porém têm uma visão mais limitada do processo, portanto, são válidos somente para as condições operacionais estudadas.

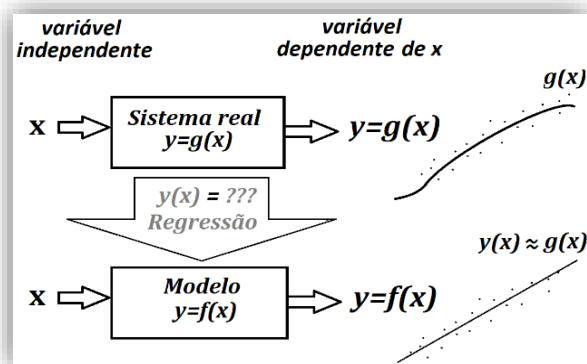


Figura I Regressão: método estatístico de estimar um modelo

A calibração consiste na comparação entre os valores obtidos pelo instrumento de medição e outro de referência, também chamado de padrão. Os resultados são registrados em um certificado, que é um documento que registra e comprova que a calibração foi realizada. No processo de calibração são testadas amostras de valores conhecidos, chamados de "calibradores" ou "dados de treinamento" de um modelo. Portanto, esse modelo, (curva de calibração), é "treinado" para estimar e reduzir imperfeições do processo real de medição executado pelo instrumento. O processo em essência "ensina" o instrumento a produzir resultados que são mais precisos do que aqueles que ocorreriam se não fosse calibrado, portanto, é uma forma de inteligência artificial. A calibração se baseia na análise de regressão [Frost], que é um método estatístico que busca identificar a relação entre uma (ou mais) variável independente e uma variável dependente, e apresenta diversas aplicações importantes nos campos da ciência e engenharia moderna [1]. A análise de regressão pertence ao do campo de *machine learning* (aprendizagem de máquina) e é usada principalmente com dois propósitos: **predição e previsão**.

2. Modelo de regressão e a calibração

2.1 Determinação do modelo

A análise de regressão [1] é apropriada para estimar funções de calibração porque podemos tomar as concentrações como valores fixos em uma variável independente x e as respostas do instrumento de medição em uma variável dependente (y). Imagine que se pretende obter a concentração exata e precisa de uma dada substância. Podemos utilizar algumas amostras de MRCs dessa substância para estimar a resposta do sensor. Mas como estimar o valor verdadeiro das demais concentrações? A solução é usar a curva de calibração desse medidor. A calibração [Mermet] envolve duas etapas: primeiro medir intensidades de sinal usando amostras de soluções com concentrações conhecidas (padrões) para construir uma curva de calibração, e em seguida podemos estimar concentrações medindo a resposta do sensor. Em uma calibração de um medidor de concentração, x_i seria o valor da concentração da amostra conhecida, por exemplo um determinado MRC, e y_i seria o valor medido da resposta à saída

do instrumento. Geralmente denomina-se a saída de 'variável de resultado' e as uma ou mais variáveis independentes à entrada de "preditores" ou "características".

2.2 Orientações para a calibração

Na calibração, deseja-se calibrar um equipamento de medição comparando-o a um padrão conhecido. Na área química, em geral a comparação é realizada usando medições a partir de um MRC. Como exemplo, podemos citar os materiais de referência certificados internacionais (ISO GUIDE 35:2006 [a]). Como nos medidores em geral a resposta é diretamente proporcional à amostra, geralmente utilizamos a regressão linear para estimar a incerteza de medição do instrumento devida à curva de calibração. Segundo o documento orientativo do Inmetro sobre validação de métodos analíticos (DOQ-CGCRE-008 [b]), o método é mais sensível quando pequenas variações de concentração resultam em uma maior variação da resposta (no caso da regressão linear, essa sensibilidade é definida pelo coeficiente angular b_1) e normalmente são necessários vários níveis de concentração (mínimo cinco) para construir a curva de calibração, e o número de réplicas em cada nível de concentração deve ser o mais próximo possível daquele empregado na rotina do laboratório.

2.3 O método dos mínimos quadrados ordinários

Um dos métodos estatísticos para estimar os coeficientes de regressão é o *Ordinary Least Squares* (OLS), ou método dos mínimos quadrados ordinários. Como exemplo (Fig. 2), considere o modelo linear simples com uma entrada x e uma saída y onde $y = \beta_0 + \beta_1 x + \varepsilon$, sendo que β_0 e β_1 são os parâmetros desconhecidos do modelo a ser estimado e ε é o erro. A estimativa de y é feita a partir $y = b_0 + b_1 x$, sendo \hat{y} , b_0 e b_1 as estimativas de y , β_0 e β_1 , e onde $e_i = y_i - \hat{y}_i$ é o resíduo.

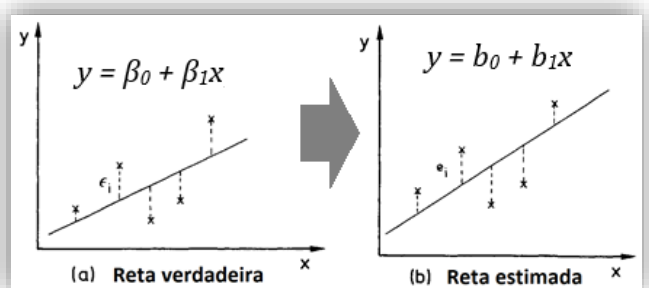


Figura 2 Reta verdadeira e reta estimada.

2.5 Tipos de regressão realizadas pelo INCAL

O INCAL realiza cinco tipos de regressão (Tabela 1).

Tabela 1 Modelos de Regressão suportados pelo INCAL

Tipo	Modelo
Linear	$y = b_0 + b_1 x$
Lin. Ponderada	$y = b_0 + b_1 x$ (pesos w_i)
Quadrática	$y = b_0 + b_1 x + b_2 x^2$
Cúbica	$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3$
Quártica	$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + b_4 x^4$
Quíntica	$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + b_4 x^4 + b_5 x^5$

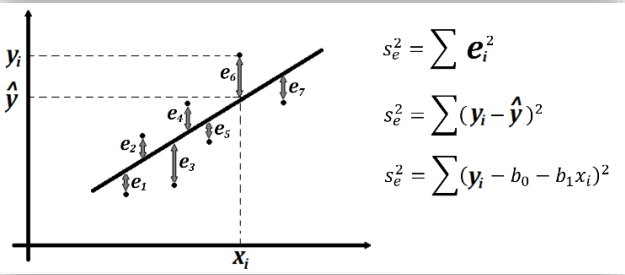


Figura 3 OLS minimiza a soma dos desvios quadrados.

Minimizando a soma quadrática dos erros, ou resíduos, é possível obter b_0 e b_1 , ou seja, a reta que apresenta a soma mínima dos quadrados dos desvios. A estimativa dos mínimos quadrados pode ser estendida aos modelos mais gerais, por exemplo, com regressão múltipla e por polinômios ou exponenciais.

2.4 INCAL: um software para cálculo da regressão

Apesar de existirem ferramentas como o Excel® para realização dos cálculos da calibração, existe ainda o trabalho referente a organizar valores, obter parâmetros, apresentar gráficos e resultados e etc. Esse trabalho pode ser facilitado com automação. O INCAL é um software desenvolvido pelo INMETRO utilizando o Visual Basic for Applications (VBA) para automatizar Excel® por meio de macros. A interface do INCAL dispõe de três setores (Fig 4): um painel de controle para seleção de modelo, um de apresentação dos resultados e outro para os gráficos.

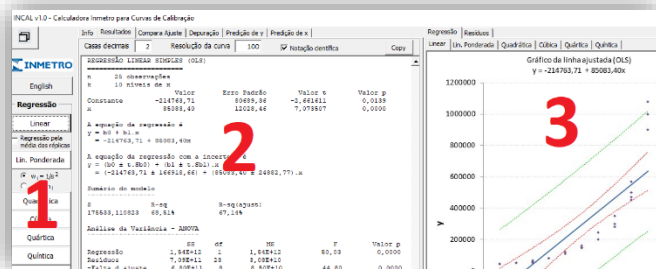


Figura 4 Setores principais da interface do INCAL.

Na Figura 5 podemos ver dois exemplos de ajuste por meio de regressão. O ajuste linear e o quadrático. É possível observar também as bandas de confiança e predição em 95%.

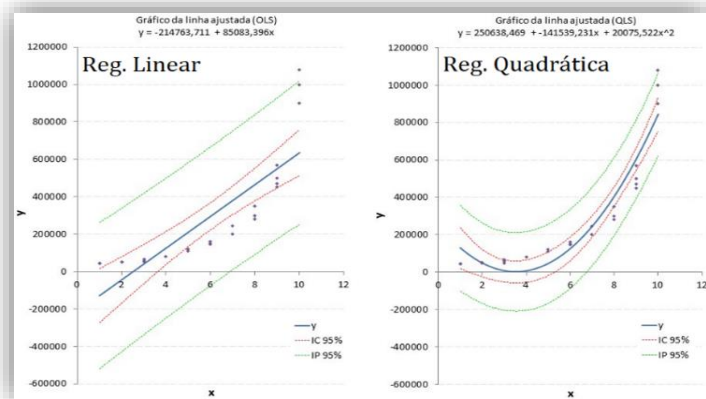


Figura 5 Exemplos de ajuste com as regressões do INCAL.

2.6 O problema do overfitting

Na calibração temos interesse em prever um valor de concentração x_0 dado a observação y_0 . Comparando a regressão linear e quártica (Fig. 6), vemos que a quártica produz um sobre ajuste do modelo. Apesar do erro padrão ter diminuído e o R^2 aumentado, o erro do intervalo de confiança de predição em y para $x_0 = 10$ praticamente dobrou, logo o modelo é mais impreciso.

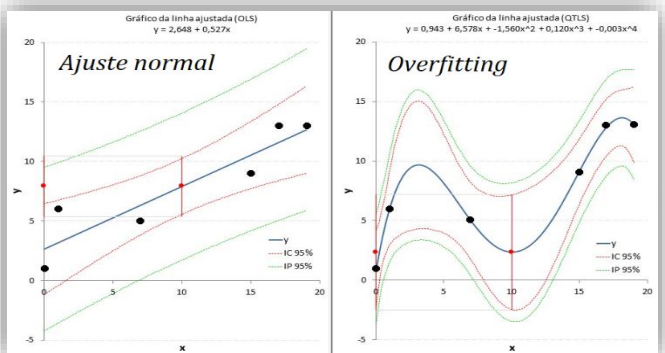


Figura 6 Predição de y com um modelo quártico (overfitting).

3. Interpretação dos resultados da calibração

3.1 Calibração de soluções de fluorescência

Considere-se [1] os dados de calibração de manganês (Mn) de um espectrômetro de emissão atômica de plasma indutivamente acoplado e que produz como resposta intensidades em unidades arbitrárias. Efetuamos a regressão linear no INCAL e os resultados são apresentados na Figura 7.

Tabela 2 Dados de uma calibração de Mn.

Concentração de Mn [ppb]	0	2	4	6	8	10
Resposta	114	870	2087	3353	3970	4950

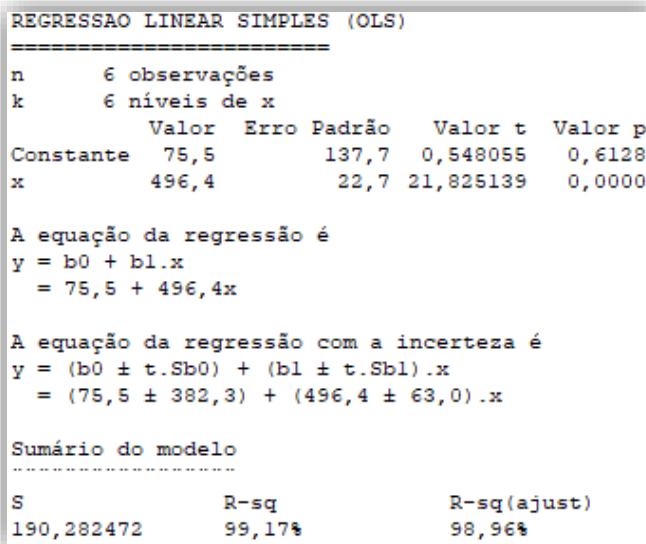


Figura 7 Resultado da regressão linear de calibração de Mn.

3.2 Predição da concentração x_0 em função de y_0

Podemos utilizar a equação da curva de calibração para prever concentrações dentro do domínio de calibração (no caso de 0 a 10 ppb). Esse seria o procedimento usual de um instrumento de medição para determinar concentrações desconhecidas. Se realizarmos várias medições da resposta, a predição é dada pela média das respostas. Com base na curva de calibração e incerteza calculadas, uma solução que apresente uma resposta de 3000 unidades iria indicar uma concentração de manganês de 5,9 $\mu\text{g/l}$ e um intervalo de predição com uma confiança de 95% entre 4,7 e 7,0 $\mu\text{g/l}$ (Fig. 8). Tendo determinado os

coeficientes do modelo, podemos também observar os testes estatísticos, teste t e o de valor p. O valor p calcula a significância do coeficiente, ou seja, a probabilidade deste ser nulo.

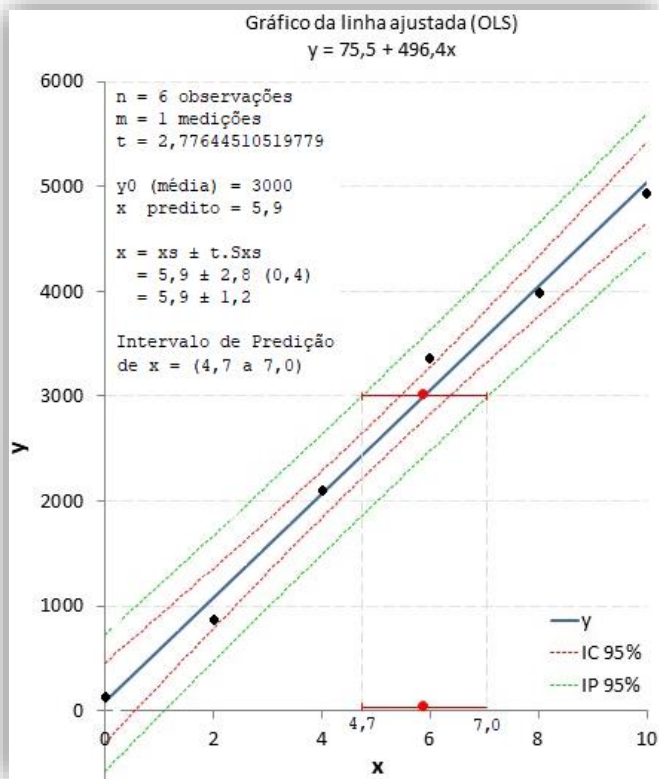


Figura 8 Intervalos de predição de x_0 para uma resposta $y_0=3000$.

3.3 Análise dos resíduos

Um gráfico residual plota os resíduos no eixo y versus os valores previstos da variável dependente no eixo x. O perfil desejado é o *imparcial* (não tendencioso) e *homocedástico*, no qual os resíduos apresentam uma distribuição normal ao longo da curva, a média dos resíduos é nula e o desvio padrão é constante. Temos exemplos de alguns perfis típicos de distribuição dos resíduos na Figura 9. Se a faixa de resíduos se afina ou se alarga ao longo da curva diz-se que é *heterocedástico*. Quanto mais estreita é a faixa dos resíduos, menor é o desvio padrão. Se o modelo é imparcial, não se deve confiar nos resultados. Em um ajuste imparcial os resíduos caem aleatoriamente em torno de zero e seguem uma distribuição normal. Podemos analisar a seguir os parâmetros de adequação do sumário do modelo (Fig. 7).

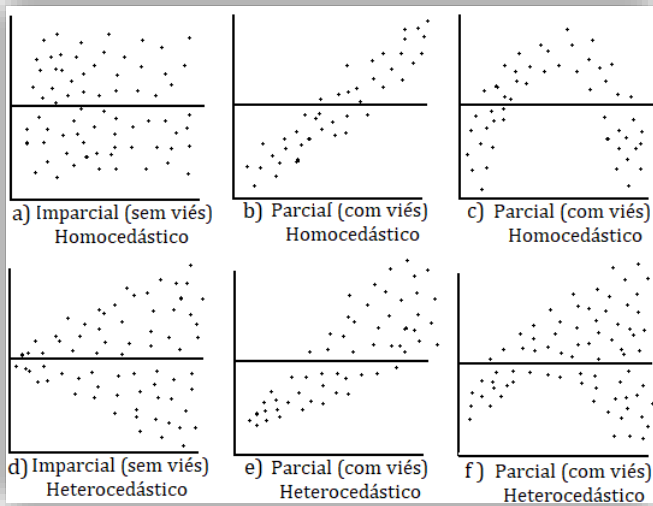


Figura 9 Perfis típicos de resíduos e sua caracterização.

3.4 Parâmetros de adequação do sumário R^2 e S

O R^2 (“*Goodness of Fit*”) é uma medida de adequação para modelos de regressão linear que indica a porcentagem de variação da variável dependente (y) que é explicada pela variação da variável independente (x) segundo o modelo. Portanto, R^2 mede a proporção da relação entre o modelo e a variável dependente e varia de 0 a 100%. R^2 não serve para determinar se as estimativas e previsões dos coeficientes são tendenciosas. Nesse caso, deve-se avaliar os gráficos residuais. O erro padrão da regressão S fornece a medida absoluta da distância típica a que os pontos de dados ficam da linha de regressão. É uma medida da dispersão das medições em torno da linha de regressão ajustada, conseqüentemente, representa a variação na resposta que não pode ser explicada pelo modelo, neste caso a reta da regressão. Como esse erro se reflete na parcela da variância que permanece inexplicada pelo modelo (equação) depois que x foi levado em consideração, também é chamado de erro padrão de y dado x , (símbolo S_y). A variância dos resíduos é calculada com o somatório dos quadrados dos resíduos dividido por $(n-2)$ e o erro padrão dos resíduos é dado pela sua raiz quadrada.

$$S = \sqrt{\frac{\sum e_i^2}{n-2}} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$$

Conforme o grau do polinômio aumenta, o número de graus de liberdade dos resíduos se reduz pois passamos a ter um número maior coeficientes a serem estimados. O erro padrão (S) e o R^2 (ou R -sq) são duas medidas-chave da adequação do ajuste do modelo. Embora o R^2 seja o mais conhecido, em geral a grande preferência por ele como indicador da qualidade do ajuste é um pouco exagerada. S e R^2 fornecem uma avaliação numérica de quão bem um modelo se ajusta aos dados de amostra, porém seus efeitos são contrapostos. Conforme os pontos de dados se movem para mais perto da linha de regressão, R^2 aumenta enquanto S diminui, e vice-versa.

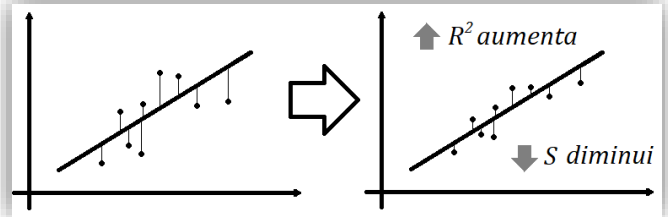


Figura 10 Efeitos em R^2 e S devidos à redução da dispersão.

4. Análise da Variância - ANOVA

Na regressão [1], a variação entre os valores de y pode ser separada em duas componentes: da regressão e dos resíduos. Isso nos permite comparar a magnitude relativa dessas componentes e efetuar análises e extrair informações sobre o sucesso da regressão. A variância dos valores em y é dada pela expressão normal da variância $\sum (y_i - \bar{y})^2 / (n - 1)$, a variância dos resíduos é dada por $\sum (y_i - \hat{y}_i)^2 / (n - 2)$, e a variância devida à regressão, ou seja, dos valores ajustados \hat{y}_i em torno da média \bar{y} é $\sum (\hat{y}_i - \bar{y})^2$. A tabela ANOVA, que é um acrônimo de “*ANALYSIS OF VARIANCE*” é então montada com base nesses graus de liberdade, somas dos quadrados e médias quadráticas.

Tabela 3 Tabela ANOVA.

Fonte de variação	G.L	Somas dos quadrados	Média Quadrática	F
Regressão	1	$\sum (\hat{y}_i - \bar{y})^2$	$\frac{\sum (\hat{y}_i - \bar{y})^2}{1}$	$\frac{MQR}{MQE}$
Resíduos (Erro)	n-2	$\sum (y_i - \hat{y}_i)^2$	$\frac{\sum (y_i - \hat{y}_i)^2}{(n-2)}$	
Total	n-1	$\sum (y_i - \bar{y})^2$	$\frac{\sum (y_i - \bar{y})^2}{(n-1)}$	

O componente “Resíduos” também pode aparecer como “Erro” na tabela ANOVA. Assim como o valor p, o valor de F pode ser utilizado como outro teste de significância do relacionamento entre y e x.

Análise da Variância - ANOVA					
	SS	df	MS	F	Valor p
Regressão	17246922	1	17246922	476	0,0000
Resíduos	144830	4	36207		
TOTAL	17391751	5			
Obs:					
- Variação explicada pelo modelo	: 99,17%				
- Variação máxima explicável	: 100,00%				

Figura II Tabela ANOVA do Calibração de fluorescência.

Voltando ao exemplo de calibração de manganês, obtivemos a tabela ANOVA (Figura II) e podemos observar o valor de $R^2 = 99,17\%$. Um valor assim elevado é bastante usual em procedimentos de análise de calibração, a não ser que algum modelo bastante inadequado esteja sendo utilizado. O valor de F elevado e a probabilidade praticamente nula no valor p da regressão são típicos de processos de calibração e não apresentam nenhum interesse específico de análise.

5. Conclusões

Apresentamos o processo de calibração por meio da regressão como um método de inteligência artificial necessário para determinação de modelos de ajuste aos dados de treinamento baseados em padrões conhecidos. Foram apresentados critérios para a determinação do modelo, orientações para a calibração e um software desenvolvido no Inmetro para automatizar os cálculos necessários e facilitar a apresentação dos resultados. A interpretação adequada dos resultados da regressão é um aspecto crucial do processo de calibração, assim como a análise dos resíduos e da variância e esses temas foram discutidos. O aspecto importante do desenvolvimento de ferramentas como o INCAL é que retrata a determinação do INMETRO de fomentar cada vez mais a cultura de inovação tecnológica e de domínio dessas tecnologias, priorizando o aumento da produtividade das pesquisas em um órgão técnico, de excelência e que é um motor do desenvolvimento do nosso parque tecnológico e industrial e da sociedade como um todo.

6. Referências

Livros

- [1] THOMPSON, M.; LOWTHIAN, P. *Notes on Statistics and Data Quality for Analytical Chemists*, 1st edition, Londres: Imperial College Press, 2011
- [2] OTTO, M. *Chemometrics: Statistics and Computer Application in Analytical Chemistry*, 2nd edition, Weinheim: Wiley-vch, 2016
- [3] MILLER, J.; MILLER, J. *Statistics and Chemometrics for Analytical Chemistry*, 6th edition, New Jersey: Prentice Hall, 2010
- [4] CHATTERJEE, S.; HADI, A. *Regression Analysis by Example*, 4th edition, Londres: Wiley, 2006
- [5] CHATTERJEE, S.; SIMON, J. *Handbook of Regression Analysis*, 1st edition, Londres: Wiley, 2013
- [6] CHRISTENSEN, R. *Analysis of Variance, Design, and Regression: Applied Statistical Methods*, 1st edition, Londres: Chapman and Hall/CRC Press, 1996
- [7] MONTGOMERY, D.; PECK, E.; Vining, G. *Introduction to Linear Regression Analysis*, 5th edition, New Jersey: Wiley, 2012

Internet

- [Frost] <https://statisticsbyjim.com/regression/>
- [Jost] <https://condor.depaul.edu/sjost/it223/documents>
- [Portal Action] <http://www.portalaction.com.br/analise-de-regressao/110-curva-de-calibracao>
- [Garret] <https://www.techtudo.com.br/dicas-e-tutoriais/noticia/2016/03/o-que-e-e-como-usar-vba-no-excel.html>
- [Ninoska] http://www.professores.uff.br/ninoska/wp-content/uploads/sites/57/2017/10/Control_Aula14_Met-Identif_2sem2017.pdf

Documentos Orientativos

- [a] ISO GUIDE 35:2006, Reference materials — General and statistical principles for certification.
- [b] Orientação sobre validação de métodos analíticos, DOQ-CGCRE-008, Revisão 09 – JUN/2020.

Créditos das Imagens

Capa

<https://br.freepik.com/fotos/icone> - Ícone foto criado por rawpixel.com

Sumário

<https://www.gratispng.com/png-w83d6l/>

<https://www.gratispng.com/png-o8k6mi/>

<https://www.gratispng.com/png-k5gel5/>

<https://www.pngegg.com/ko/png-wnplg/>

<https://www.gratispng.com/png-3tu8ut/>

<https://www.pngitem.com/>

[middle/hRJJRow_transparent-machine-learning-png-machine-learning-images-png/](https://www.pngitem.com/middle/hRJJRow_transparent-machine-learning-png-machine-learning-images-png/)

Artigos

p.02

<https://br.freepik.com/fotos/coracao> - Coração foto criado por onlyyouqj

p.03

<https://nexxto.com/iot-na-saude-a-transformacao-do-setor/>

p.10

<https://br.freepik.com/fotos/fundo> - Fundo foto criado por kjpargeter

p.16

<https://br.freepik.com/vetores/fundo> - Fundo vetor criado por starline

p.27

<https://br.freepik.com/vetores/tecnologia> - Tecnologia vetor criado por rawpixel.com

p.28

<https://br.freepik.com/vetores/fundo> - Fundo vetor criado por rawpixel.com

p.29

<https://br.freepik.com/vetores/abstrato> - Abstrato vetor criado por upklyak

p.30

<https://www.shutterstock.com/pt/image-vector/big-data-artificial-intelligence-concept-machine-1092234560>

p.40

<https://br.freepik.com/vetores/bandeira>>Bandeira vetor criado por starline