

**UNIVERSIDADE DE TAUBATÉ**  
**Maria Clara de Moura Santos Coelho Ferreira**

**CARACTERIZAÇÃO BIOQUÍMICA DA SALIVA DE  
PACIENTES FUMANTES E NÃO FUMANTES UTILIZANDO  
ESPECTROSCOPIA FT-IR E TÉCNICAS DE MACHINE  
LEARNING**

**Taubaté-SP**  
**2021**

**UNIVERSIDADE DE TAUBATÉ**  
**Maria Clara de Moura Santos Coelho Ferreira**

**CARACTERIZAÇÃO BIOQUÍMICA DA SALIVA DE  
PACIENTES FUMANTES E NÃO FUMANTES UTILIZANDO  
ESPECTROSCOPIA FT-IR E TÉCNICAS DE MACHINE  
LEARNING**

Trabalho de Conclusão de Curso  
apresentado ao Departamento de  
Odontologia da Universidade de  
Taubaté como parte dos requisitos  
para obtenção do título de  
bacharel em Odontologia  
Área de Concentração:  
Estomatologia  
Orientador: Prof. Dr. Luis Felipe  
das Chagas e Silva de Carvalho

**Taubaté-SP**  
**2021**

**Grupo Especial de Tratamento da Informação - GETI**  
**Sistema Integrado de Bibliotecas – SIBi**  
**Universidade de Taubaté - UNITAU**

F383c Ferreira, Maria Clara de Moura Santos Coelho  
Caracterização bioquímica da saliva de pacientes fumantes e não  
fumantes utilizando espectroscopia FT-IR e técnicas de *Machine Learning* /  
Maria Clara de Moura Santos Coelho Ferreira. -- 2021.  
34 f. : il.

Monografia (graduação) – Universidade de Taubaté, Departamento de  
Odontologia, Taubaté, 2021.

Orientação: Prof. Dr. Luis Felipe das Chagas e Silva de Carvalho,  
Departamento de Odontologia.

1. Biofluidos. 2. Biópsia óptica. 3. Espectroscopia FT-IR. 4. Fumantes. 5.  
Saliva. I. Universidade de Taubaté. Departamento de Odontologia. III. Título.

CDD – 616.994

**Maria Clara de Moura Santos Coelho Ferreira**

**CARACTERIZAÇÃO BIOQUÍMICA DA SALIVA DE  
PACIENTES FUMANTES E NÃO FUMANTES UTILIZANDO  
ESPECTROSCOPIA FT-IR E TÉCNICAS DE MACHINE  
LEARNING**

Trabalho de Graduação,  
apresentado ao Departamento de  
Odontologia da Universidade de  
Taubaté como parte dos requisitos  
para obtenção do título de  
bacharel em Odontologia

Área de Concentração:  
Estomatologia

Orientador: Prof. Dr. Luis Felipe  
das Chagas e Silva de Carvalho

Data: 09/12/2021

Resultado: \_\_\_\_\_

**BANCA EXAMINADORA**

Prof. Dr. Luís Felipe das Chagas e Silva de Carvalho / Universidade de Taubaté

Assinatura \_\_\_\_\_

Prof. Dr. Marcelo Goncalves Cardoso / Universidade de Taubaté

Assinatura \_\_\_\_\_

Prof. Me. Odalicio Vieira de Siqueira / Universidade de Taubaté

Assinatura \_\_\_\_\_

Suplente

Prof. Me. Alexandre Cursino de Moura Santos / Universidade de Taubaté

Assinatura \_\_\_\_\_

## **AGRADECIMENTOS**

Agradeço primeiramente a Deus, por me guiar com sabedoria durante minha jornada, por me dar forças e não me deixar desistir.

A minha família, especialmente meus pais, Marcelo e Angélica e minha irmã Ana Clara pelo apoio incondicional, pelo incentivo durante essa jornada e por todo esforço investido em minha educação.

Aos professores do Departamento de Odontologia por toda dedicação para com os alunos.

Ao meu orientador, Prof. Dr. Luis Felipe das Chagas e Silva de Carvalho, pela atenção, confiança e por me orientar com sabedoria.

Aos funcionários do Departamento de Odontologia pela atenção e dedicação diária com os alunos e a instituição.

Ao CNPq pela bolsa de estudos (processo nº 154852/2020-1)

Aos meus amigos, colegas e a todos que direta ou indiretamente me ajudaram a realizar esse sonho, muito obrigada.

Dedico esse trabalho à Deus por nunca me deixar desistir e aos meus pais e à minha irmã por todo suporte e incentivo durante esses anos!

Essa vitória também é de vocês!

É necessário se espantar, se indignar e se contagiar, só assim é possível mudar a realidade...

**Nise da Silveira**

## RESUMO

Dentre as neoplasias malignas orais, mais de 90% dos casos são de carcinoma espinocelular, originadas do epitélio da mucosa oral com etiologia multifatorial extrínsecas e/ou intrínsecas. Infelizmente, detecção de câncer no estágio inicial (carcinoma in situ) ainda é um desafio. Nesse contexto clínico, surge a espectroscopia de absorção no infravermelho (Espectroscopia FT-IR). Atualmente, a espectroscopia FT-IR está sendo usada na área médica como uma técnica não invasiva de diagnóstico exigindo uma pequena quantidade amostral de tecido e biofluidos. Obtendo o resultado de forma rápida, além de auxiliar nos exames histopatológicos, sorológicos e bioquímicos. Esse procedimento pode ser usado para diversos fins como diagnosticar e monitorar doenças sistêmicas, monitorar o esforço físico de atletas e diagnosticar o câncer. O presente estudo objetiva-se comparar amostras de saliva de pacientes fumantes e não fumantes por espectroscopia FT-IR. Os voluntários desse estudo foram divididos em três grupos: controle, fumante, e fumante esporádico. As amostras foram analisadas pela região de "fingerprint" abrangendo a frequência de  $900\text{ cm}^{-1}$  a  $1800\text{ cm}^{-1}$ , pois a maior parte das informações bioquímicas são encontradas nesta região. No grupo controle analisamos onze amostras e obtivemos trinta e três espectros, no grupo fumante esporádico analisamos oito amostras e obtivemos vinte e quatro espectros e no grupo fumante analisamos nove amostras e obtivemos vinte e sete espectros. Como conclusão, a espectroscopia se apresenta como uma técnica não invasiva, de fácil instrumentação, com baixo custo financeiro, segura para o profissional e para o paciente, e eficiente para análise de biofluidos. Nesse trabalho, a técnica apresentou ótima efetividade para separar os pacientes dos grupos fumantes e não fumantes, evidenciando diferenças espectrais, principalmente, nas bandas de DNA, Proteínas, Colágeno, Amida I e Amida II.

**Palavras-chave:** Espectroscopia FT-IR; Biofluidos; Biopsia Óptica; Saliva; Fumantes



## ABSTRACT

Within the oral malignant neoplasms, more than 90% of the cases are squamous cell carcinoma, originating from the oral mucosa epithelium with extrinsic and/or intrinsic multifactorial etiology. Unfortunately, detection of early stage cancer (carcinoma in situ) still a challenge. In this clinical context, emerges infrared absorption spectroscopy (FT-IR spectroscopy). Currently, FT-IR spectroscopy is being used in the medical field as a non-invasive diagnostic technique. Obtaining the result quickly and helping in histopathological, serological and biochemical exams. This procedure can be used for various purposes such as diagnosing and monitoring systemic diseases, monitoring the physical effort of athletes and diagnosing cancer. The aim of this present study is comparing the saliva between smokers and non smokers by vibrational FT-IR spectroscopy. The volunteers in this study were divided into three groups: control, smoker, and occasional smoker. Samples were analyzed by the "fingerprint" region covering the frequency of  $900\text{ cm}^{-1}$  to  $1800\text{ cm}^{-1}$ , because more biochemical information are found in this region. In the control group we analyzed eleven and got thirty-three spectra, in the sporadic smoker group we analyzed it and got twenty-four spectra and in the smoker group we analyzed nine and got twenty-seven spectra. In conclusion, spectroscopy presents itself as a non-invasive technique, easy to instrument, with low financial cost, safe for the professional and for the patient, and efficient for biofluid analysis. In this study, the technique presents excellent effectiveness to separate patients from the smoker and non-smoker groups, showing spectral differences, mainly, in the DNA, Proteins, Collagen, Amide I and Amide II bands.

**Keywords:** FT-IR spectroscopy; Biofluids; Optical Biopsy; Saliva; Smokers

## **LISTA DE FIGURAS**

Figura 1- Coletor Universal Estéril

Figura 2- Microtubo tipo Eppendorf

Figura 3 - Espectrômetro Thermo Scientific Nicolet 6700 ATR FT-IR

Figura 4 - Espectro médio e desvio padrão de cada grupo

Figura 5 - Espectro médio de cada grupo

Figura 6- Comparação entre os modelos estatísticos

Figura 7- Confusion Matrix

Figura 8 - Comparação entre os modelos estatísticos para o grupo fumante contra o grupo controle

Figura 9 - Confusion Matrix para o grupo fumante contra o grupo controle

Figura 10 - Confusion Matrix para o grupo fumante esporádico contra o grupo controle

Figura 11- Comparação entre os modelos estatísticos para o grupo fumante esporádico contra o grupo controle

Figura 12 - Confusion Matrix para o grupo fumante contra o grupo fumante esporádico

Figura 13 - Comparação entre os modelos estatísticos para o grupo fumante contra o grupo fumante esporádico

## **LISTA DE TABELAS**

Tabela 1- Modos vibracionais referentes aos picos espectrais encontrados no grupo controle

## **LISTA DE ABREVIATURAS E SIGLAS**

ATR – *Attenuated Total Reflection* / Reflexão Total Atenuada

FT – *Fourier Transform* / Transformada de Fourier

IA – Inteligência Artificial

IR – *Infrared* / Infravermelho

ML – *Machine Learning* / Aprendizado de Máquinas

TCLE – Termo de Consentimento Livre e Esclarecido

INCA – Instituto Nacional de Câncer

# LISTA DE ANEXOS

## TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

### Dados de identificação

Título do Projeto: Análise da saliva por espectroscopia FT-IR

Pesquisador Responsável: Maria Clara de Moura Santos Coelho Ferreira, Anna Laura Almeida Barreto

Equipe executora: Maria Clara de Moura Santos Coelho Ferreira, Anna Laura Almeida Barreto, Luis Felipe das Chagas e Silva de Carvalho

Instituição a que pertence o Pesquisador Responsável: Universidade de Taubaté

Telefones para contato: (12) 981016687 (12)988682511

Nome do voluntário: \_\_\_\_\_

Idade: \_\_\_\_\_

R.G. \_\_\_\_\_

O (A) Sr. (ª) está sendo convidado(a) a participar do projeto de pesquisa tendo como título “Análise da saliva por espectroscopia FT-IR”, de responsabilidade do pesquisador Luis Felipe das Chagas e Silva de Carvalho.

A pesquisa se justifica para que seja desenvolvida uma forma de diagnóstico rápido e eficaz permitindo que as doenças bucais sejam diagnosticadas em tempo real. A maior importância reside no diagnóstico de biomarcadores na saliva, pois objetivamos avaliar a saliva de pacientes com diversas comorbidades. O procedimento em questão será realizado através da coleta de saliva por expectoração. Após as análises iremos observar se existem marcadores nesta saliva e quais são estes, o paciente receberá um adequado tratamento pelo próprio cirurgião-dentista responsável.

Após obtenção dos dados, os resultados serão comparados com a análise laboratorial. A aquisição dos dados só será realizada uma vez no paciente não havendo a necessidade de realizar novas coletas. Você será esclarecido (a) sobre a pesquisa em qualquer aspecto que desejar. Você é livre para recusar-se a participar, retirar seu consentimento ou interromper a participação a qualquer momento. A sua participação é voluntária e a recusa em participar

não irá acarretar qualquer penalidade ou perda de benefícios. Os pesquisadores irão tratar a sua identidade com padrões profissionais de sigilo. Os resultados da pesquisa serão enviados para você e permanecerão confidenciais. Seu nome ou o material que indique a sua participação não será liberado sem a sua permissão, não sendo utilizado para outras finalidades que não sejam relacionadas a pesquisa. Você não será identificado (a) em nenhuma publicação que possa resultar deste estudo. A participação no estudo não acarretará custos para você e não será disponível nenhuma compensação financeira adicional, uma vez que esta pesquisa não acarretará nenhum dano e o procedimento cirúrgico já seria realizado, mesmo sem a pesquisa. A qualquer momento você poderá solicitar novas informações e motivar sua decisão se assim o desejar. Em caso de dúvidas, acessar os responsáveis pela pesquisa Dr. Luis Felipe das Chagas e Silva de Carvalho, pelo telefone (12) 981016687 ou pessoalmente no Departamento de Odontologia da Universidade de Taubaté, situado na Rua dos Operários, 09, Centro, Taubaté.

Eu, \_\_\_\_\_, RG nº \_\_\_\_\_  
fui informado (a) e concordo em participar, como voluntário, do projeto de pesquisa acima descrito.

Taubaté, \_\_\_\_ de \_\_\_\_\_ de \_\_\_\_\_

\_\_\_\_\_  
Nome e assinatura do paciente

\_\_\_\_\_  
Nome e assinatura do pesquisador

\_\_\_\_\_  
Testemunha

\_\_\_\_\_  
Testemunha

## SUMÁRIO

<b>1 INTRODUÇÃO E JUSTIFICATIVA</b>	<b>15</b>
<b>2 PROPOSIÇÃO</b>	<b>18</b>
<b>3 REVISÃO DE LITERATURA</b>	<b>19</b>
<b>4 METODOLOGIA</b>	<b>22</b>
<b>5 RESULTADOS</b>	<b>25</b>
<b>6 DISCUSSÃO</b>	<b>31</b>
<b>7 CONCLUSÃO</b>	<b>33</b>
<b>REFERÊNCIAS</b>	<b>34</b>

## 1 INTRODUÇÃO E JUSTIFICATIVA

A neoplasia, segundo Brasileiro Filho (2016), pode ser conceituada como:

lesão constituída por proliferação celular anormal, descontrolada e autônoma, em geral com perda ou redução de diferenciação, em consequência de alterações em genes ou proteínas que regulam a multiplicação e a diferenciação das células.

A neoplasia é dividida em benigna e maligna. As benignas, geralmente, não trazem risco de vida para os indivíduos portadores da lesão; as malignas são lesões popularmente conhecidas como câncer que podem levar o indivíduo a óbito.

O processo de formação do câncer, chamado de carcinogênese é dividido em estágio de iniciação, promoção e progressão. O estágio de iniciação é a etapa onde as células sofrem alterações estruturais genéticas, pela ação dos agentes cancerígenos, tornando-as capaz de se multiplicar de forma neoplásica futuramente. No estágio de promoção, as células continuam a ser afetadas pelos agentes cancerígenos que as alteram ainda mais, transformando-as em células malignas. E no estágio de progressão, a neoplasia torna-se visível devido à rápida, desordenada e irreversível multiplicação de suas células (INCA, 2019).

Dentre as neoplasias malignas orais, mais de 90% dos casos são de carcinoma espinocelular, também conhecido como carcinoma epidermóide, com etiologia multifatorial extrínsecas como fumo e/ou intrínsecas como a deficiência nutricional. O risco de câncer bucal aumenta em três vezes para os indivíduos fumantes em comparação aos indivíduos não fumantes, além disso, o risco também aumenta pelo tempo e pela relação dose-dependente. Além de apresentar mais de setenta substâncias carcinogênicas como nitrosamina e benzeno, o fumo do tabaco representa aproximadamente 80% dos casos de carcinoma espinocelular. (NEVILLE et al., 2013).

De acordo com Freitas et al., (2016, p.16):

A associação entre o fumo e o álcool na etiologia do câncer de boca está relacionada com esse efeito de aumento da permeabilidade causado pelo álcool, que pode ocasionar um aumento da penetração dos carcinógenos presentes no tabaco.

Clinicamente, a lesão do carcinoma epidermóide pode se apresentar como exofítica (aumento de volume), endolítica (aspecto invasivo e/ou ulcerado), leuplástica (mancha esbranquiçada), eritoplástica (mancha avermelhada), leucoeritoplástica (machas avermelhadas e esbranquiçadas) e dificilmente o paciente apresentará sintomatologia dolorosa. Em casos mais avançados da



doença ocorre a destruição do osso subjacente com sintomatologia dolorosa variável para cada paciente e na radiografia é observado uma imagem radiolúcida em forma de “ruídos de traça” com bordas irregulares (NEVILLE et al., 2013).

A principal forma de diagnóstico do câncer são os exames histopatológicos, que detectam alterações celulares a nível morfológico. Porém essa técnica apresenta algumas desvantagens como a contaminação ambiental pelos agentes utilizados para preparar as amostras, o tempo demasiado para obter o resultado e a capacidade profissional que o patologista deve possuir. Por isso, a razão pela qual as técnicas de biopsia ópticas, em especial a espectroscopia FT-IR, vem sendo cada vez mais estudada é pela detecção de mudanças teciduais a níveis bioquímicos. A espectroscopia FT-IR tem o potencial de ser utilizada como uma técnica para o diagnóstico, detecção, discriminação e progressão de doenças pela análise da composição química da célula (BUNACIU ET AL, 2015).

A espectroscopia vibracional é um método em desenvolvimento para o diagnóstico precoce de doenças. Essa técnica interessou diversos pesquisadores pela capacidade de caracterização de biomoléculas, pela facilidade de manusear os instrumentos, pela análise rápida e não invasiva. Por meio da análise espectral de uma determinada amostra, observa-se os níveis de lipídeos, proteínas, ácidos nucleicos e carboidratos, dessa forma, os dados obtidos podem ser utilizados para distinguir amostras saudáveis das patológicas.(CAMERON et al., 2018).

A espectroscopia vibracional pode ser utilizada para analisar alterações nas biomoléculas de tecidos saudáveis e patológicos, e com isso caracterizar os mesmo. Uma das vantagens da técnica além da análise não invasiva de biofluidos e a pequena quantidade necessária para ser examinada, aproximadamente 1ml, a espectroscopia emite informações sobre a área de “fingerprint”, região onde se encontram grande parte das moléculas biológicas como carboidratos, lipídeos, ácidos nucleicos e proteínas que ajudam na caracterização dos tecidos. (LEAL et al., 2018).

Nesse contexto clínico, a Inteligência Artificial recebe destaque. De acordo com Braga et. al., (2019, p. 16408) :

A inteligência artificial (IA) é um campo da ciência da computação que imita os processos de pensamento humano, a capacidade de aprendizagem e o armazenamento de conhecimento. Atualmente a IA tem sido fortemente utilizada na análise de imagens e previsões, vários ramos da medicina têm desenvolvido técnicas de aperfeiçoamento da

inteligência artificial vinculada a melhoria da saúde no seu conceito mais global aceito.

Na área da saúde, existem diversos dados que podem ser obtidos dos pacientes desde clínicos, como exames laboratoriais e de imagens, até dados genômicos com as informações contidas no DNA e dados proteômicos dados contidos nos conjuntos de proteínas presentes em uma amostra biológica. Com isso, o ML que consiste em uma área da IA que tem o objetivo de prever resultados, nesse caso diagnóstico e prognósticos mediante técnicas estatísticas e dados fornece uma técnica para reconhecer padrões de dados massivos e usar algoritmos para prever o resultado futuro dos pacientes (Dhillon e Singh, 2019).

Este trabalho tem o intuito de caracterizar as amostras de saliva de pacientes fumantes e não fumantes pela espectroscopia FT-IR, mostrar a efetividade da técnica para o diagnóstico precoce e monitoramento do câncer bucal e para a utilização como exame complementar ao histopatológico. Ademais, este trabalho tem a finalidade de explicar os biomarcadores encontrados nas amostras espectrais.

## **2 PROPOSIÇÃO**

O presente trabalho tem por objetivos comparar amostras de saliva de pacientes fumantes e não fumantes por espectroscopia vibracional FT-IR associada ao aprendizado de máquina; contribuir para o estabelecimento da técnica de biopsia óptica pela espectroscopia FT-IR; investigar alterações intrínsecas a nível molecular na saliva, investigar a sensibilidade e especificidade do teste para a identificação de pacientes fumantes e não fumantes por meio da saliva pelos seus biomarcadores. Esta pesquisa tem o intuito de contribuir com pesquisas futuras para mostrar a efetividade da técnica para o diagnóstico precoce e monitoramento do câncer bucal e para a utilização como exame complementar ao histopatológico.

### 3 REVISÃO DE LITERATURA

O câncer, de maneira geral, atinge um número expressivo da população mundial, devido ao aumento dos indivíduos a exposição de fatores carcinogênicos além disso possui alta taxa de mortalidade, principalmente, se detectado tardiamente. Dessa forma, existe uma grande preocupação para enfrentar a doença com o diagnóstico precoce, já que permite maior chances de cura, novas modalidades de tratamento e medidas preventivas (Brasileiro Filho, 2016) (Freitas et. al., 2016). O carcinoma epidermóide, também conhecido como carcinoma de células escamosas ou carcinoma espinocelular, representa mais de 90% dos tumores malignos da região oral (Freitas et. al., 2016) (Neville et. al., 2016).

A formação do câncer é conhecida como carcinogênese e envolve o processo de iniciação, promoção e progressão. Na fase de iniciação, as células sofrem alteração no DNA pelos agentes carcinogênicos; na fase de progressão, elas se alteram de maneira mais drástica e se transforma em células malignas e na fase de progressão a neoplasia torna-se visível devido à rápida, desordenada e irreversível multiplicação de suas células (Freitas et. al, 2016) e (INCA, 2019).

Existem vários fatores etiológicos para o câncer bucal sendo o tabagismo, o álcool e a exposição solar considerados os principais. O fumo em si representa 80% dos casos de carcinoma epidermóide já que contêm mais de setenta substâncias carcinogênicas (Neville et al., 2016). Clinicamente as lesões se apresentam como úlceras ou nódulos, ou podem até mesmo ser precedidas de lesões potencialmente cancerizáveis que são as leucoplasias, as eritroplasias, e a queilite actínica (Lemos Junior et al, 2013).

A principal forma de diagnóstico do câncer são os exames histopatológicos, que detectam alterações celulares a nível morfológico. Porém essa técnica apresenta algumas desvantagens como a contaminação ambiental pelos agentes utilizados para preparar as amostras, o tempo demasiado para obter o resultado e a capacidade profissional que o patologista deve possuir (Bunaciu et. al., 2015) e (Bunaciu et. al., 2014).

Dessa forma a biopsia óptica, em especial a espectroscopia no infravermelho por transformada de Fourier (FT-IR), vem sendo estudada cada vez

mais pela detecção de mudanças teciduais a níveis bioquímicos. A espectroscopia FT-IR tem o potencial de ser utilizada como uma técnica para o diagnóstico, detecção, discriminação e progressão de doenças pela análise da composição química da célula (Bunaciu et. al., 2015). A técnica está sendo utilizada como alternativa aos exames laboratoriais, em função da alta sensibilidade para analisar tecido cardíaco, tecido esquelético, e fluidos corporais como sangue, saliva e urina. (Caetano Júnior et. al., 2015).

A técnica consiste na emissão de luz policromática na matéria e permite medir a quantidade de luz absorvida por uma amostra na presença da luz infravermelha e com os dados coletados utiliza-se a Transformada de Fourier para converter os dados brutos em espectros. (Naseer et. al., 2020). As vantagens da técnica consistem na fácil instrumentação, na mínima ou nenhuma preparação amostral, na pouca quantidade amostral e no rápido diagnóstico in vitro (Bunaciu et. al., 2015) e (Naseer et. at., 2020).

Essa técnica pode ser usada na rotina clínica, pois não é prejudicial à saúde e requer pequena quantidade amostral, fornecendo uma impressão digital da amostra que contém a estrutura, o conteúdo e a possível modificação bioquímica da mesma (Ukkonen et al., 2019; Cameron et.al., 2018; Finlayson, et. al. , 2019). Ela pode ser chamada de biopsia óptica pela sua capacidade de analisar as propriedades ópticas das amostras evidenciando suas alterações bioquímicas e moleculares (Carvalho, 2008 e Carvalho, 2011)

A saliva tem a possibilidade de monitorar alterações corporais e mediante ela é possível realizar o controle de doping e observar as alterações fisiológicas em atletas, devido ao treinamento físico. Além disso apresenta vantagens em relação ao sangue com menor risco de contaminação e a coleta não é invasiva. (Caetano Júnior et. al., 2015).

Além da facilidade de coleta, os biofluidos, apresentam diversos componentes que podem ser considerados e utilizados como biomarcadores para detectar e observar a progressão de diversas doenças. A saliva em especial, pode ser usada para diagnosticar e monitorar algumas doenças sistêmicas afetam as glândulas salivares e a composição da saliva como Síndrome de Sjögren, cirrose alcoólica, fibrose, diabetes mellitus e doenças das glândulas suprarrenais (Leal et. al., 2018).

Com isso, a IA ganha destaque pois permite aperfeiçoar o diagnóstico, prognóstico e o tratamento em várias áreas da medicina e dessa forma mostra os seus vários benefícios para a saúde prevendo com mais precisão a evolução da doença, no desempenho dos tratamentos (Braga et. al., 2019). De acordo com Dhillon e Singh, 2019, o ML tem sido proposto na área da saúde pois permite compreender e melhorar os dados sobre a saúde do paciente e melhora na tomada de decisões.

Nos estudos feitos por Rizzato em 2019 e Magalhães em 2020, as autoras propuseram a avaliar treze amostras de saliva de pacientes fumantes e não fumantes por espectroscopia FT-IR a fim de encontrar biomarcadores de estruturas bioquímicas e moleculares. Rizzato em 2019, concluiu que por conta da obtenção de diferentes bandas pela análise de ambos os grupos é preciso estender este estudo com o objetivo de observar um grupo com câncer bucal e comprovar um possível biomarcador para o diagnóstico precoce de câncer de boca.

Magalhães em 2020 conclui que, a espectroscopia FT-IR, é capaz de fornecer informações preciosas para a diferenciação das amostras de saliva entre pacientes fumantes e não fumantes e que os modos vibracionais do grupo amido, proteínas, fosfato e ácidos nucleicos podem ser considerados possíveis marcadores moleculares para demonstrar tal diferença.

## 4 METODOLOGIA

Este trabalho foi aprovado pelo Comitê de Ética em Pesquisa, da Universidade de Taubaté sob o número 19436919.7.0000.5501. Todos os voluntários são maiores de 18 anos, sem distinção de sexo e antes das coletas de amostras de saliva serem realizadas, esses voluntários assinaram o Termo de Consentimento Livre e Esclarecido (TCLE), sendo que as amostras foram coletadas no Departamento de Odontologia da Universidade de Taubaté e em uma clínica particular em Joinville. Os voluntários foram divididos em três grupos: controle, fumante, e fumante esporádico.

No grupo controle, foram coletadas e analisadas amostras de saliva de onze voluntários que nunca experimentaram qualquer forma de tabaco. Desses onze voluntários, um apresentava alergia e refluxo e fazia uso contínuo de Pantoprazol e Dexatametasona, um apresentava colesterol alto e fazia uso contínuo de Rosuvastatina e um apresentava diabetes e pressão alta fazendo uso contínuo diário de Rosuvastatina, Glifage, Glimpirida e Losartana. Outros dois voluntários faziam uso contínuo e diário de anticoncepcional, um de Tamisa 30 e outro de Belara. No grupo fumante, foram coletadas e analisadas nove amostras de pacientes com mais de 40 anos que fumam mais de um maço de cigarro por dia.

No grupo fumante esporádico, foram coletadas e analisadas 8 amostras de pacientes que fumam esporádica e/ou socialmente. Deste grupo um paciente fuma exclusivamente cigarro de palha, um voluntário fuma cannabis, um voluntário já experimentou tabaco e cannabis, um voluntário experimentou maconha, cigarro de palha e tabaco, um voluntário experimentou cannabis e fuma narguilé, um voluntário experimentou cigarro de palha, um voluntário fuma narguilé, e um voluntário fuma cigarro e cannabis.

As amostras de saliva foram coletadas por cuspe/expectoração. Os participantes da pesquisa foram instruídos a bochechar água durante um minuto e, permanecer sem deglutir por alguns minutos e, ao final cuspir toda a saliva em um coletor universal estéril. Todas as amostras foram levadas ao vórtex para que fosse homogeneizadas. Depois de realizado esse processo as amostras foram transferidas do coletor universal estéril para o micro tubo tipo Eppendorf com ajuda de uma pipeta calibrada em 1000µm. As amostras foram armazenadas em um

freezer em  $-20^{\circ}\text{C}$  no Centro de Pesquisas Odontológicas (CEPEO), localizado no departamento de Odontologia da Universidade de Taubaté, em Taubaté -SP. As amostras foram transferidas em um isopor com gelo para o Instituto de Pesquisa Energéticas e Nucleares (Ipen) no Centro de Laser e Aplicações (CLA) no dia pré-determinado para a análise.



Figura 1 - Coletor Universal Estéril



Figura 2 - Microtubo tipo Eppendorf

O espectrômetro utilizado foi o Espectrômetro Thermo Scientific Nicolet 6700 ATR FT-IR que possui cristal de diamante para aquisição dos dados espectrais. No equipamento as amostras foram inseridas sem aditivos sob o cristal com o auxílio de uma pipeta calibrada em  $1\ \mu\text{l}$  para obtenção dos espectros. Depois de colocadas no cristal, esperou-se pela secagem completa das amostras para obtenção de melhores espectros. O tempo de secagem variou de 2 a 7 minutos com tempo médio de 5 minutos. A análise foi feita com 32 scans de varredura para obtenção da média dos espectros com resolução de  $4\text{cm}^{-1}$ . Os espectros foram obtidos em triplicata.

Após cada análise, as amostras foram removidas do cristal com o auxílio de papel absorvente e limpo com álcool 92,8% para retirada total das amostras do cristal, evitando contaminação entre amostras. Foram realizados 32 scans de background do equipamento a cada amostra analisada para retirar as possíveis interferências.

Os espectros foram pré-processados para se tornarem adequados para a comparação na etapa de análise. Alguns artefatos podem ser incluídos nos espectros durante sua aquisição, e nesta etapa eles serão minimizados sem prejuízo as amostras coletadas. Para a realização do estudo, o pré-processamento



e a obtenção dos espectros foram feitos pelo software Origin Pro8.5 e pelo software Orange e com a ajuda do artigo realizado por Movasaghi, et al., 2018 foi evidenciado os picos bioquímicos.



Figura 3 - Espectrômetro Thermo Scientific Nicolet 6700 ATR FT-IR

## 5 RESULTADOS

As amostras foram analisadas pela região de “fingerprint” abrangendo a frequência de  $900\text{ cm}^{-1}$  a  $1800\text{ cm}^{-1}$  pois nessa região concentra-se mais informações bioquímicas sobre as amostras. No grupo controle analisamos onze amostras e obtivemos trinta e três espectros, no grupo fumante esporádico analisamos oito amostras e obtivemos vinte e quatro espectros e no grupo fumante analisamos nove amostras e obtivemos vinte e sete espectros. Na imagem abaixo observamos a média espectral de cada grupo e o desvio padrão deles, sendo o grupo controle correspondente a cor azul, o grupo fumante esporádico, à cor vermelha e o grupo fumante, à cor verde.

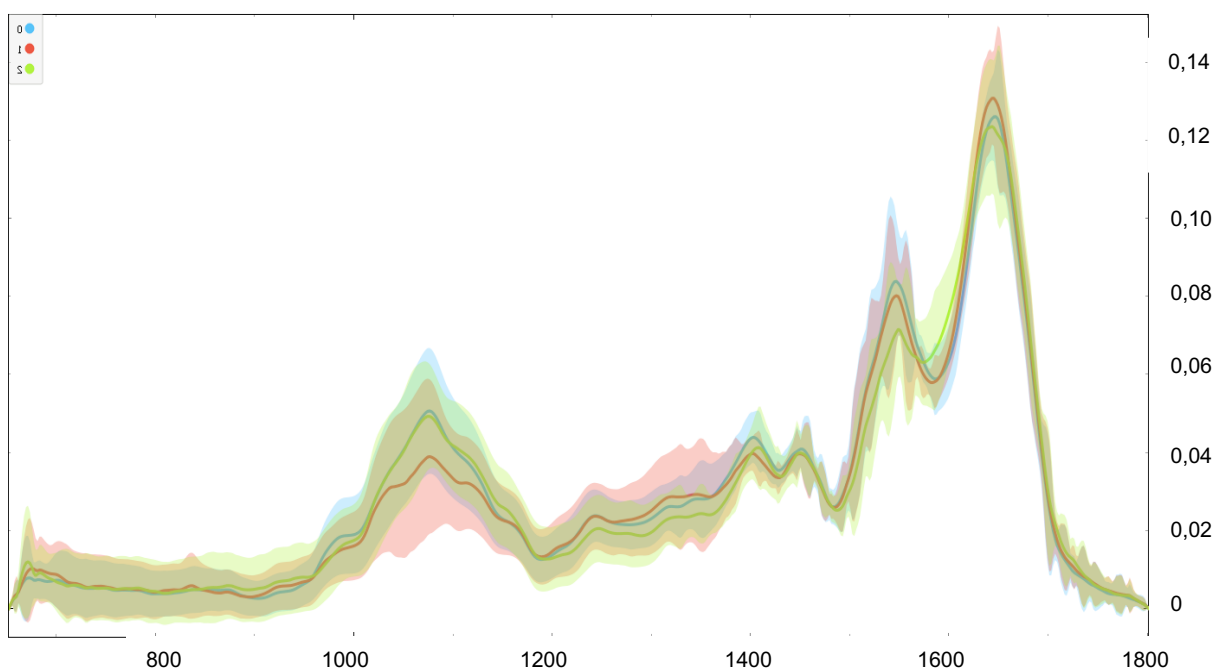


Figura 4 - Espectro médio e desvio padrão de cada grupo

Foram evidenciados os picos  $1076\text{ cm}^{-1}$  que corresponde à conformação cis esquelética de DNA e alongamento simétrico de fosfato  $[\text{PO}_2]$ ,  $1403\text{ cm}^{-1}$  que corresponde à modos simétricos de  $\text{CH}_3$  do grupos metil de proteínas e  $\text{CH}_3$  de colágeno,  $1451\text{ cm}^{-1}$  que corresponde à modos de flexão assimétricos de  $\text{CH}_3$  do grupos metil de proteínas,  $1547\text{ cm}^{-1}$  corresponde aproximadamente à banda de proteína, amida II, peptídeo amida II e amida II de proteínas e  $1646\text{ cm}^{-1}$  que corresponde à amida I, citosina  $\text{C}_5$  metilada, ligação  $\text{C}=\text{O}$ , alongamento  $\text{C}=\text{C}$  uracil e  $\text{NH}_2$  guanina.

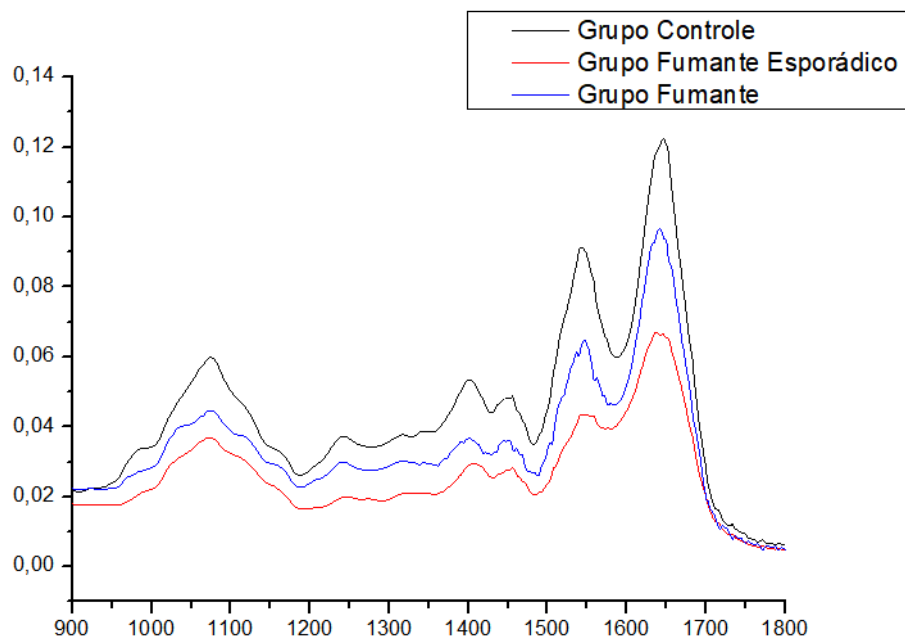


Figura 5 - Espectro médio de cada grupo

<i>Bandas (cm-1)</i>	<i>Modos Vibracionais</i>	<i>Componentes estruturais</i>
1076	Alongamento simétrico de fosfato	DNA
1403	CH <sub>3</sub> dos grupos metil e CH <sub>3</sub>	Proteínas e Colágeno
1451	CH <sub>3</sub> dos grupos metil	Proteínas
1547		Proteína, amida II, peptídeo amida II e amida II de proteínas
1646	Ligação alongamento C=C NH <sub>2</sub> guanina	C=O, Amida I, citosina C <sub>5</sub> uracil e metilada

Tabela 1- Modos vibracionais referentes aos picos espectrais encontrados no grupo controle

A fim de verificar quanto o software consegue prever as semelhanças e diferenças entre as amostras, foi utilizada a técnica “Leave one out” que consistem em uma análise supervisionada treinada previamente sobre um conjunto de dados pré-definidos, classificando novos dados de acordo com os critérios definidos pelo treinamento.

Nesse processo utilizamos modelos estatísticos que servem para informar a especificidade (recall) e sensibilidade (precision) sendo eles “kNN”, “Tree”, “SVM”, “SGD,” “Random Forest”, “Neural Network”, “Naive Bayes”, “Logistic Regression”, “Gradient Boosting” e “AdaBoost”. O melhor modelo precisa ser aquele que se aproxime de 1 e ao compararmos os três grupos simultaneamente, o modelo que se mostrou mais eficiente foi a “Neural Network” com 0,715 e 0,713 para sensibilidade e especificidade, respectivamente. Na “Confusion Matrix”, observamos quantas amostras seriam separadas corretamente para seus respectivos grupos sendo 72,7% para o grupo controle, 65,5% para fumante esporádico e 75% para fumante.

Evaluation Results					
Model	AUC	CA	F1	Precision	Recall
kNN	0.717	0.584	0.575	0.588	0.584
Tree	0.754	0.634	0.633	0.632	0.634
SVM	0.783	0.634	0.612	0.638	0.634
SGD	0.722	0.644	0.635	0.633	0.644
Random Forest	0.800	0.634	0.621	0.627	0.634
Neural Network	0.857	0.713	0.712	0.715	0.713
Naive Bayes	0.676	0.545	0.535	0.536	0.545
Logistic Regression	0.641	0.475	0.379	0.397	0.475
Gradient Boosting	0.819	0.663	0.657	0.668	0.663
AdaBoost	0.692	0.604	0.601	0.605	0.604

Figura 6- Comparação entre os modelos estatísticos dos três grupos

		Predicted		
		0	1	2
Actual	0	72.7 %	20.7 %	10.7 %
	1	9.1 %	65.5 %	14.3 %
	2	18.2 %	13.8 %	75.0 %

Figura 7- Confusion Matrix dos três grupos

O teste “Leave one out” foi repetido com o intuito de comparar o grupo controle e fumante, fumante e fumante esporádico e controle e fumante esporádico. Nesse processo utilizamos os mesmos modelos estatísticos do teste anterior. Quando comparado o grupo fumante ao controle, o modelo que se mostrou mais eficiente foi o “SGD” com 0,853 para sensibilidade e especificidade. Na “Confusion Matrix”, observamos quantas amostras seriam separadas corretamente para seus respectivos grupos sendo 86% para o grupo controle, 84,4% para fumantes.

Evaluation Results					
Model	AUC	CA	F1	Precision	Recall
kNN	0.768	0.720	0.714	0.722	0.720
Tree	0.748	0.733	0.734	0.739	0.733
SVM	0.839	0.787	0.785	0.787	0.787
SGD	0.850	0.853	0.853	0.853	0.853
Random Forest	0.819	0.760	0.759	0.759	0.760
Neural Network	0.894	0.840	0.837	0.846	0.840
Naive Bayes	0.737	0.680	0.680	0.680	0.680
Logistic Regression	0.657	0.560	0.424	0.535	0.560
Gradient Boosting	0.827	0.760	0.759	0.759	0.760
AdaBoost	0.760	0.760	0.761	0.762	0.760

Figura 8 - Comparação entre os modelos estatísticos para o grupo fumante contra o grupo controle

		Predicted	
		0	2
Actual	0	86.0 %	15.6 %
	2	14.0 %	84.4 %

Figura 9 - Confusion Matrix para o grupo fumante contra o grupo controle

Ao comparar o grupo fumante esporádico com o controle, o modelo que se mostrou mais eficiente foi o “SGD” com 0,803 e 0,809 para sensibilidade e especificidade, respectivamente. Na “Confusion Matrix”, observamos quantas amostras seriam separadas corretamente para seus respectivos grupos sendo 76,7% para o grupo controle e 73,9% para o grupo fumante esporádico.

		Predicted	
		0	1
Actual	0	76.7 %	26.1 %
	1	23.3 %	73.9 %

Figura 10 - Confusion Matrix para o grupo fumante esporádico contra o grupo controle

Evaluation Results					
Model	AUC	CA	F1	Precision	Recall
kNN	0.752	0.712	0.714	0.725	0.712
Tree	0.592	0.576	0.576	0.576	0.576
SVM	0.858	0.788	0.786	0.787	0.788
SGD	0.805	0.803	0.804	0.809	0.803
Random Forest	0.842	0.758	0.754	0.756	0.758
Neural Network	0.850	0.773	0.775	0.785	0.773
Naive Bayes	0.570	0.515	0.519	0.537	0.515
Logistic Regression	0.429	0.591	0.439	0.349	0.591
Gradient Boosting	0.782	0.773	0.773	0.774	0.773
AdaBoost	0.742	0.742	0.744	0.749	0.742

Figura 11- Comparação entre os modelos estatísticos para o grupo fumante esporádico contra o grupo controle

Quando comparado o grupo fumante ao fumante esporádico, o modelo que se mostrou mais eficiente foi o “Neural Network” com 0,787 e 0,788 para sensibilidade e especificidade. Na “Confusion Matrix”, observamos quantas amostras seriam separadas corretamente para seus respectivos grupos sendo 71,4% para o grupo controle, 78,8% para fumantes.

		Predicted	
		1	2
Actual	1	71.4 %	21.2 %
	2	28.6 %	78.8 %

Figura 12 - Confusion Matrix para o grupo fumante contra o grupo fumante esporádico

Evaluation Results					
Model	AUC	CA	F1	Precision	Recall
kNN	0.704	0.738	0.738	0.753	0.738
Tree	0.588	0.623	0.624	0.625	0.623
SVM	0.797	0.738	0.738	0.738	0.738
SGD	0.719	0.721	0.722	0.723	0.721
Random Forest	0.818	0.754	0.754	0.755	0.754
Neural Network	0.854	0.787	0.787	0.788	0.787
Naive Bayes	0.713	0.705	0.705	0.705	0.705
Logistic Regression	0.563	0.557	0.399	0.311	0.557
Gradient Boosting	0.794	0.721	0.722	0.727	0.721
AdaBoost	0.738	0.738	0.738	0.741	0.738

Figura 13 - Comparação entre os modelos estatísticos para o grupo fumante contra o grupo fumante esporádico

## 6 DISCUSSÃO

No Brasil, o câncer é a segunda maior causa de mortes depois das doenças cardiovasculares e o número de pessoas acometidas pela doença vem crescendo. Isso se deve a rotina estressante e agitada que a população leva, recorrendo às bebidas alcoólicas, ao tabaco e aos *fast foods* para aliviar essa rotina e proporcionar uma forma de relaxamento. Outro fator que proporcionou uma parte do aumento de casos, foi a pandemia pelo Covid 19 pois além da apreensão da população para ir a hospitais e clínicas para a realização de exames houve atrasos na obtenção deles.

Quando se trata do câncer bucal, o Instituto Nacional de Câncer (INCA) estimou para 2020 mais de onze mil casos para a população masculina brasileira, representando 5% de todos os tipos de câncer. É imprescindível nesses casos não pensar em biópsia, exame “padrão ouro” para a complementação do diagnóstico clínico (Carvalho, 2008). Apesar disso, a biópsia apresenta algumas desvantagens como a comorbidade gerada ao paciente, o pós-operatório que exige atenção, o conhecimento e o tempo, de aproximadamente sete dias, para se obter esses resultados.

Dessa forma, se faz necessário outros exames que contribuam ao diagnóstico, mas que permitam o conforto do paciente. As pesquisas vêm concretizando que a espectroscopia FT-IR é uma técnica segura ao paciente e ao profissional além da facilidade para analisar biofluidos que consegue observar as características bioquímicas da amostra por meio de suas propriedades ópticas (Finlayson, et. al., 2019 e Carvalho, 2008).

É importante ressaltar que grande parte da literatura sobre câncer bucal estima que aproximadamente 80% dos casos da doença se devem ao uso prolongado de tabaco, com ou sem fumaça. Ao analisar o cigarro, já que é a forma de fumo mais comum e recorrente, pesquisadores evidenciaram a presença de mais de setenta substâncias cancerígenas. Dada a importância desses dados faz-se justificável as análises do presente estudo.

Outrossim, os resultados obtidos com esse trabalho são condizentes aos esperados visto que comprovam as alterações bioquímicas devido ao uso prolongado e até esporádico de tabaco. Apesar dos achados bioquímicos nas



bandas de DNA, Proteínas, Colágeno, Amida I e Amida II serem possíveis biomarcadores para o câncer bucal, é precipitado dizer que a espectroscopia visa substituir os exames histopatológicos pois apesar das suas vantagens essa técnica não tem o intuito de substituir e sim complementar para que o diagnóstico seja mais preciso.

Nesse contexto, a IA, em especial o Machine Learning se destaca como uma ferramenta para auxiliar a espectroscopia e evidenciar sua especificidade e precisão. Nesse estudo conseguimos observar a eficácia da técnica pelas ferramentas de ML com o teste “Leave one out” e pela tabela da “Confusion Matrix”.

É preciso associar as técnicas de IA à Espectroscopia FTIR pois permite que os dados possam ser analisados em um software previamente configurado e treinado. Dessa forma é evidente a necessidade de maiores estudos para consolidar a técnica de biopsia óptica e apesar de suas vantagens serem superiores ao exame histopatológico, seu intuito não é de substituição, mas sim de complementação para um diagnóstico mais rápido e mais preciso.

## **7 CONCLUSÃO**

Mediante o exposto, a espectroscopia se apresenta como uma técnica não invasiva, de fácil instrumentação, com baixo custo financeiro, segura para o profissional e para o paciente, e eficiente para análise de biofluidos. Nesse trabalho, a técnica apresentou ótima efetividade para separar os pacientes dos grupos fumantes e não fumantes, evidenciando diferenças espectrais, principalmente, nas bandas de DNA, Proteínas, Colágeno, Amida I e Amida II. Dessa forma, os achados bioquímicos são possíveis biomarcadores para diferenciar os grupos fumantes e não fumantes. Apesar dos resultados preliminares, faz-se necessários outros estudos com maiores números amostrais para consolidar a técnica na clínica médica-odontológica.

## REFERÊNCIAS

- 1 Brasileiro Filho G. Bogliolo Patologia. 9<sup>a</sup>. ed. Rio de Janeiro: Guanabara Koogan; 2016
- 2 INCA. Como surge o câncer? 2021; Aug. 04. Disponível em: <https://www.inca.gov.br/como-surge-o-cancer>. Acesso em: 17 Aug. 2021.
- 3 Neville BW, et al. Patologia oral e maxilofacial. 4<sup>a</sup>. ed. Rio de Janeiro: Elsevier; 2016.
- 4 Freitas RM, et al. Fatores de risco e principais alterações citopatológicas do câncer bucal: uma revisão de literatura. Revista Brasileira de Análises Clínicas. 2016. 48(1): 13-18. Disponível em: <http://www.rbac.org.br/artigos/fatores-de-risco-e-principais-alteracoes-citopatologicas-do-cancer-bucal-uma-revisao-de-literatura/>. Acesso em: 21 ago. 2021.
- 5 Bunaciu AA, et al. Applications of FT-IR Spectrophotometry in Cancer Diagnostics. Critical Reviews in Analytical Chemistry. 2015. 45(2): 156-165. Disponível em: [https://www.researchgate.net/publication/270648928\\_Applications\\_of\\_FT-IR\\_Spectrophotometry\\_in\\_Cancer\\_Diagnostics](https://www.researchgate.net/publication/270648928_Applications_of_FT-IR_Spectrophotometry_in_Cancer_Diagnostics). Acesso em 14 ago. 2021
- 6 Cameron JM, et al. Biofluid Spectroscopic Disease Diagnostics: A Review on the Processes and Spectral Impact of Drying. Journal of Biophotonics. 2018. 11(4): 1-12. Disponível em: [https://www.researchgate.net/publication/322781334\\_Biofluid\\_Spectroscopic\\_Disease\\_Diagnostics\\_A\\_Review\\_on\\_the\\_Processes\\_and\\_Spectral\\_Impact\\_of\\_Drying](https://www.researchgate.net/publication/322781334_Biofluid_Spectroscopic_Disease_Diagnostics_A_Review_on_the_Processes_and_Spectral_Impact_of_Drying). Acesso em 18 ago. 2021.
- 7 Leal LB, et al. Vibration spectroscopy and body biofluids: Literature review for clinical applications. Photodiagnosis and Photodynamic Therapy. 2018. 17: 237–244.
- 8 Braga AV, et al. Machine learning: O Uso da Inteligência Artificial na Medicina. Brazilian Journal of Development. 2019. 27:16407-16413.
- 9 Dhillon A, Singh A. Machine Learning in Healthcare Data Analysis: A Survey. Journal of Biology and Today's World. 2019. 21:1-10.
- 10 Lemos Junior CA. Câncer de boca baseado em evidências científicas. Associação Paulista de Cirurgiões-Dentistas. 2013. Jul 25: 178-186.
- 11 Bunaciu AA, et al. Cancer Diagnosis by FT-IR Spectrophotometry. Revue Roumaine de Chimie. 2014. 10:415-426.
- 12 Caetano Júnior PC, et al. Analysis of saliva by Fourier transform infrared spectroscopy for diagnosis of physiological stress in athletes. Research on Biomedical Engineering. 2015. 31(2):116-124.

- 13** Naseer K, et al. ATR-FTIR spectroscopy as the future of diagnostics: a systematic review of the approach using biofluids. *Applied Spectroscopy Reviews*. 2020. 56: 1-13. Disponível em: [https://www.researchgate.net/publication/340012344\\_ATR-FTIR\\_spectroscopy\\_as\\_the\\_future\\_of\\_diagnostics\\_a\\_systematic\\_review\\_of\\_the\\_approach\\_using\\_bio-fluids](https://www.researchgate.net/publication/340012344_ATR-FTIR_spectroscopy_as_the_future_of_diagnostics_a_systematic_review_of_the_approach_using_bio-fluids). Acesso em: 15 ago. 2021.
- 14** Ukkonen H, et al. Biochemical Changes in Irradiated Oral Mucosa: A FTIR Spectroscopic Study. *Biosensors*. 2019. 9(1): 12. Disponível em: <https://www.mdpi.com/2079-6374/9/1/12/htm>. Acesso em: 15 jul. 2021.
- 15** Finlayson D, et al. Is Infrared Spectroscopy Ready for the Clinic? *Analytical Chemistry*. 2019. 10(91):12117-12128. Disponível em: [https://www.researchgate.net/publication/335735637\\_Is\\_Infrared\\_Spectroscopy\\_Ready\\_for\\_the\\_Clinic](https://www.researchgate.net/publication/335735637_Is_Infrared_Spectroscopy_Ready_for_the_Clinic). Acesso em: 19 sep. 2021
- 16** Carvalho LFCS. Análise ex vivo de hiperplasia fibrosa inflamatória de Mucosa jugal por espectroscopia FT-Raman [Dissertação]. São José dos Campos: Faculdade de Odontologia de São José dos Campos, Universidade Estadual Paulista; 2008.
- 17** Carvalho LFCS. Análise de queilite actínica por espectroscopia micro FT-IR [Tese]. São José dos Campos: Faculdade de Odontologia, Campus de São José dos Campos, UNESP; 2011.
- 18** Rizzato JMB. Caracterização da saliva de pacientes fumantes por espectroscopia FT-IR [Trabalho de Conclusão de Curso]. Taubaté: Departamento de Odontologia da Universidade de Taubaté; 2019.
- 19** Magalhães NC. Análise de biomarcadores na saliva de pacientes fumantes e não fumantes por espectroscopia FT-IR – estudo piloto [Dissertação]. Taubaté: Departamento de Odontologia da Universidade de Taubaté; 2020.
- 20** Movasaghi Z, et al. Fourier Transform Infrared (FTIR) Spectroscopy of Biological Tissues. *Applied Spectroscopy Reviews*. 2008. 43: 134-179. Disponível em: <http://dx.doi.org/10.1080/05704920701829043>. Acesso em 09. Jun. 2021

Autorizo a reprodução e divulgação total ou parcial desta obra por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citadas as fontes.

Maria Clara de Moura Santos Coelho Ferreira

Taubaté, dezembro de 2021