



Universidade do Minho  
Escola de Engenharia

André Filipe Gonçalves Névoa Fernandes Braga

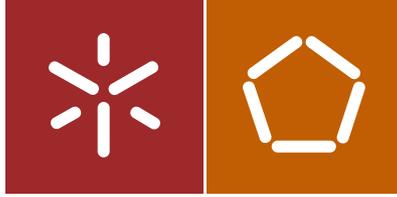
*Pervasive Patient Timeline*

André Filipe Gonçalves Névoa Fernandes Braga *Pervasive Patient Timeline*

UMinho | 2015

outubro de 2015





Universidade do Minho  
Escola de Engenharia

André Filipe Gonçalves Névoa Fernandes Braga

*Pervasive Patient Timeline*

Dissertação de Mestrado  
Ciclo de Estudos Integrados Conducentes ao Grau de  
Mestre em Engenharia e Gestão de Sistemas de Informação

Trabalho efectuado sob a orientação do  
Professor Doutor Carlos Filipe da Silva Portela

e coorientação do  
Professor Doutor Manuel Filipe Vieira Torres dos Santos

## DECLARAÇÃO

Nome: André Filipe Gonçalves Névoa Fernandes Braga

Endereço eletrónico: andre.nevoabraga@gmail.com      Telefone: 939696210

Bilhete de Identidade/Cartão do Cidadão: 14199932

Título da dissertação: *Pervasive Patient Timeline*

Orientadores:

Professor Doutor Carlos Filipe da Silva Portela

Professor Doutor Manuel Filipe Vieira Torres dos Santos

Ano de conclusão: 2015

Mestrado em Engenharia e Gestão de Sistemas de Informação

DE ACORDO COM A LEGISLAÇÃO EM VIGOR, NÃO É PERMITIDA A REPRODUÇÃO DE QUALQUER PARTE DESTA TESE/TRABALHO.

Universidade do Minho, 30/10/2015

Assinatura: 

## **Agradecimentos**

O reconhecimento aqui feito é direcionado a todas as pessoas que de alguma forma contribuíram para o desenvolvimento desta dissertação e que de uma forma geral fizeram parte do meu percurso académico.

Aos meus pais e irmão, agradeço esta oportunidade de poder investir no meu futuro, todo o apoio e companhia durante todo percurso académico, que me permitiram entre altos e baixos chegar a este momento final.

Ao Professor Doutor Carlos Filipe da Silva Portela, meu Orientador, tenho a agradecer o facto de me ter aceitado como seu orientando, permitindo-me a oportunidade de participar no projeto no qual me mostrei interessado. Agradeço a sua quase omnipresença em todos os momentos em que requisitei ajuda para ultrapassar as dificuldades sentidas, as orientações e ensinamentos e o cultivo de um ambiente de trabalho fácil e agradável.

Ao Professor Doutor Manuel Filipe Vieira Torres dos Santos, meu Coorientador, agradeço o apoio e oportunidade de poder fazer parte deste projeto.

Ao Centro Hospitalar do Porto e Hospital de Santo António, quero agradecer o fornecimento dos dados facultados para o desenvolvimento deste trabalho, bem como o interesse demonstrado em melhorar os apoios a doentes e profissionais de saúde através do produto deste trabalho.

Quero agradecer também os meus amigos de curso, David Baía, Marco Rodrigues, Pedro Cunha, Pedro Matos e Rui Rocha, por me permitirem fazer parte de um grupo espetacular onde reina a anarquia do pensamento, por aceitarem de uma forma peculiar as características mais únicas de cada um e por todos os momentos bons e maus que passamos juntos, mas que fazem deles memórias lembradas com nostalgia. Um bem-haja a todos eles pela amizade, insanidade e irreverência.

Ao Eduardo Jorge, quero agradecer todo o apoio ao longo de muitos anos de percurso académico, e o facto de me continuar a ajudar a crescer como pessoa. Durante o desenvolvimento, desta dissertação tenho a agradecer-lhe a disponibilidade para me ajudar e a partilha dos seus conhecimentos de medicina.



## ***Pervasive Patient Timeline***

### **Resumo**

Em Medicina Intensiva, a apresentação de informação médica nas Unidades de Cuidados Intensivos (UCI) é feita de diversas formas (gráficos, tabelas, texto, ...), pois depende do tipo de análises realizadas, dos dados recolhidos em tempo real pelos sistemas de monitorização, entre outros. A forma como é apresentada a informação pode dificultar a leitura da condição clínica dos doentes por parte dos profissionais de saúde, principalmente quando há a necessidade de um cruzamento entre vários tipos de dados clínicos/fontes de informação.

A evolução das tecnologias para novos padrões como a ubiquidade e o *pervasive* torna possível a recolha e o armazenamento de vários tipos de informação, possibilitando um acesso em tempo-real sem restrições de espaço e tempo. A representação de *timelines* em papel transformou-se em algo desatualizado e por vezes inutilizável devido às diversas vantagens da representação em formato digital. O uso de Sistemas de Apoio à Decisão Clínica (SADC) em UCI não é uma novidade, sendo que a sua principal função é facilitar o processo de tomada de decisão dos profissionais de saúde. No entanto, a associação de *timelines* a SADC, com o intuito de melhorar a forma como a informação é apresentada, é uma abordagem inovadora, especialmente nas UCI.

Este trabalho procura explorar uma nova forma de apresentar a informação relativa aos doentes, tendo por base o espaço temporal em que os eventos ocorrem. Através do desenvolvimento de uma *Pervasive Patient Timeline* interativa, os profissionais de saúde terão acesso a um ambiente, em tempo real, onde podem consultar o historial clínico dos doentes, desde a sua admissão na unidade de cuidados intensivos até ao momento da alta. Torna-se assim possível visualizar os dados relativos a sinais vitais, análises clínicas, entre outros. A incorporação de modelos de *Data Mining* (DM) produzidos pelo sistema INTCare é também uma realidade possível, tendo neste âmbito sido induzidos modelos de DM para a previsão da toma de vasopressores, que foram incorporados na *Pervasive Patient Timeline*. Deste modo os profissionais de saúde passam assim a ter uma nova plataforma capaz de os ajudar a tomarem decisões de uma forma mais precisa.

**Palavras-Chave:** Unidades de Cuidados Intensivos, *Pervasive Patient Timeline*, INTCare, Medicina Intensiva, Sistemas de Apoio à Decisão Clínica, *Data Mining*



# Pervasive Patient Timeline

## Abstract

In Intensive Care Medicine, the presentation of medical information in the Intensive Care Units (ICU) is done in many shapes (graphics, tables, text,...). It depends on the type of exams executed, the data collected in real time by monitoring systems, among others. The way in which information is presented can make it difficult for health professionals to read the clinical condition of patients. When there is the need to cross between several types of clinical data/information sources the situation is even worse. The evolution of technologies for emerging standards such as ubiquity and pervasive makes it possible to gather and storage various types of information, thus making it available in real time and anywhere. Also with the advancement of technologies, the representation of timelines on paper turned into something outdated and sometimes unusable due to the many advantages of representation in digital format. The use of Clinical Decision Support Systems (CDSS) is not a novelty, and its main function is to facilitate the decision-making process, through predictive models, continuous information monitoring, among others. However, the association of timelines to CDSS, in order to improve the way information is presented, is an innovative approach, especially in the ICU.

This work seeks to explore a new way of presenting information about patients, based on the time frame in which events occur. By developing an interactive Pervasive Patient Timeline, health professionals will have access to an environment in real time, where they can consult the medical history of patients. The medical history will be available from the moment in which patients are admitted in the ICU until their discharge, allowing health professionals to analyze data regarding vital signs, medication, exams, among others. The incorporation of Data Mining (DM) models produced by the INTCare system is also a reality, and in this context, DM models were induced for predicting the intake of vasopressors, which were incorporated in Pervasive Patient Timeline. Thus health professionals will have a new platform that can help them to make decisions in a more accurate manner.

**Keywords:** Intensive Care Units, Pervasive Patient Timeline, INTCare, Intensive Care Medicine, Clinical Decision Support Systems, Data Mining



# Índice

Resumo.....	v
Abstract.....	vii
Lista de Figuras.....	xi
Lista de Tabelas.....	xii
Lista de Abreviaturas, Siglas e Acrónimos.....	xiii
Capítulo 1 – Enquadramento do Trabalho Proposto.....	1
1.1 Enquadramento e Motivação.....	1
1.2 Objetivos.....	2
1.3 Estrutura do Documento.....	3
Capítulo 2 – Revisão de Literatura.....	5
2.1 Estratégia de Revisão de Literatura.....	5
2.2 Medicina Intensiva e Unidades de Cuidados Intensivos.....	6
2.2 Sistemas de Apoio à Decisão.....	7
2.3 Descoberta de Conhecimento em Base de Dados.....	12
2.4 <i>Data Mining</i> .....	15
2.5 Sistema INTCare.....	20
2.6 Agentes Anti-Hipotensivos.....	22
2.7 <i>Pervasive Computing</i> .....	23
2.8 A <i>Timeline</i> .....	26
Capítulo 3 – Abordagem Metodológica.....	31
3.1 Metodologia de Investigação.....	31
3.2 Modelo do Processo de Tomada de Decisão.....	33
3.3 Metodologia de <i>Data Mining</i> .....	35
3.4 Cruzamento entre DSR e MPTD.....	36
3.5 Cruzamento entre DSR e CRISP-DM.....	38

Capítulo 4 – Desenvolvimento da Solução .....	41
4.1 Inteligência.....	41
4.2 Desenho .....	41
4.3 Escolha.....	44
4.4 Implementação .....	45
4.5 Monitorização.....	50
Capítulo 5 – Descrição do Estudo de Data Mining .....	51
5.1 Compreensão do Negócio.....	51
5.2 Compreensão dos Dados.....	52
5.3 Preparação dos Dados .....	56
5.4 Modelação .....	61
5.5 Avaliação .....	64
5.6 Implementação .....	68
Capítulo 6 – Discussão do Trabalho Realizado .....	71
Capítulo 7 – Conclusão.....	73
7.1 Análise dos Riscos.....	74
Referências .....	77
Anexos .....	82
Anexo I – Avaliação do Desempenho da <i>Pervasive Patient Timeline</i> .....	82
Anexo II – Artigos Científicos.....	84
1. Step Towards a Patient Timeline in Intensive Care Units .....	84
2. Data Mining to predict the use of Vasopressors in Intensive Medicine Patients.....	85
3. Real-Time models to predict the use of vasopressors in monitored patients .....	86

## Lista de Figuras

Figura 1 – Processo de DCBD.....	13
Figura 2 – Exemplo de curvas ROC.....	19
Figura 3 – Sistema INTCare e subsistemas.....	21
Figura 4 – A queda de Troia em <i>Chronicle</i> de Eusebius.....	27
Figura 5 – <i>Chart of Universal History</i> de Thomas Jefferys.....	28
Figura 6 – Exemplo de uma <i>timeline</i> digital.....	29
Figura 7 – Modelo de Processo da Metodologia <i>Design Science Research</i> .....	31
Figura 8 – Modelo do Processo de Tomada de Decisão.....	34
Figura 9 – Ciclo de vida do CRISP-DM.....	35
Figura 10 – Exemplo da SIMILE Timeline.....	42
Figura 11 – Exemplo da Vis.js.....	43
Figura 12 – Exemplo da Timeglider JS.....	43
Figura 13 – Representação da <i>Pervasive Patient Timeline</i> .....	49
Figura 14 – Distribuição qualitativa da variável VSPGeral.....	59
Figura 15 – Construção dos cenários sem classe.....	62
Figura 16 – Construção dos modelos com classes.....	62
Figura 17 – Melhor curva ROC da análise SCL.....	67
Figura 18 – Melhor curva ROC da análise CL1.....	67
Figura 19 – Melhor curva ROC da análise CL2.....	68
Figura 20 – Modelos DM na <i>Pervasive Patient Timeline</i> .....	69
Figura 21 – Avaliação de Desempenho da <i>Pervasive Patient Timeline</i> – Parte 1.....	82
Figura 22 – Avaliação de Desempenho da <i>Pervasive Patient Timeline</i> – Parte 2.....	83

## Lista de Tabelas

Tabela 1 – Matriz de Confusão .....	18
Tabela 2 – Relação entre área e capacidade classificativa .....	19
Tabela 3 – Cruzamento DSR vs MPTD .....	37
Tabela 4 – Cruzamento DSR vs CRISP-DM .....	38
Tabela 5 – Cruzamento das <i>timelines</i> e suas características .....	45
Tabela 6 – Variáveis e respectivas tabelas e fontes .....	54
Tabela 7 – Distribuição qualitativa das variáveis comuns às abordagens Sem Classes e Com Classes .....	55
Tabela 8 – Análise quantitativa das variáveis na abordagem Sem Classes .....	56
Tabela 9 – Representação das variáveis e valores de referência por classe .....	60
Tabela 10 – Distribuição qualitativa das variáveis na abordagem Com Classes .....	61
Tabela 11 – Cenários e grupos de variáveis .....	63
Tabela 12 – Parâmetros dos algoritmos de DM .....	64
Tabela 13 – Três melhores modelos por métrica, e por análise .....	65
Tabela 14 – Modelo de Análise de Riscos .....	75

## **Lista de Abreviaturas, Siglas e Acrónimos**

AC – Análises Clínicas

AD – Árvores de Decisão

AMD – Admissão Doente

CL1 – Com Classes 1

CL2 – Com Classes 2

CHP – Centro Hospitalar do Porto

CM – *Case Mix*

CRISP-DM – *Cross Industry Standard Process for Data Mining*

DCBD – Descoberta de Conhecimento em Base de Dados

DM – *Data Mining*

DR – *Design Research*

DSR – *Design Science Research*

EPAS – *Enterprise Patient Administration System*

ETL – *Extract, Transform and Loading*

EUA – Estados Unidos da América

FEE – Folha de Enfermagem Eletrónica

FN – Falsos Negativos

FP – Falsos Positivos

FRC – Farmácia

HELP – *Health Evaluation through Logical Processes*

HL7 – *Health Level 7*

JSON – *JavaScript Object Notation*

MLG – Modelos Lineares Generalizados

MIT – Massachusetts *Institute of Technology*

MPTD – Modelo do Processo de Tomada de Decisão

NB – Naïve Bayes

OCDE – Organização para a Cooperação e Desenvolvimento Económico

PCE – Processo Clínico Eletrónico

PHP – *Hypertext Preprocessor*

ROC – *Receiver Operating Characteristic*

SAD – Sistema de Apoio à Decisão

SADC – Sistema de Apoio à Decisão Clínica

SCL – Sem Classes

SDAL – Sistema de Dor Abdominal de Leeds

SPAD – Sistema *Pervasive* de Apoio à Decisão

SQL – *Structured Query Language*

SV – Sinais Vitais

SVM – *Support Vector Machines*

UCI – Unidade de Cuidados Intensivos

VN – Verdadeiros Negativos

VP – Verdadeiros Positivos

XAMPP – *Cross-plataform Apache, MySQL, PHP, Perl*

## **Capítulo 1 – Enquadramento do Trabalho Proposto**

### **1.1 Enquadramento e Motivação**

Nas últimas décadas, temos observado o crescimento da aplicação das tecnologias às mais diversas áreas, como forma de trazer mais e melhor conhecimento sobre a área em estudo. A medicina é uma dessas áreas e tem sido alvo de imensas transformações graças à introdução das tecnologias (Direção Geral de Saúde, 2003). Mais especificamente, os sistemas de informação na medicina procuram garantir uma maior qualidade e eficiência, não só no tratamento de doentes, mas também na gestão das instituições de saúde (Haux et al., 2004).

No âmbito da Medicina Intensiva, surgiu o projeto de investigação INTCare, implementado na Unidade de Cuidados Intensivos (UCI) do Hospital de Santo António, resultado da parceria conjunta entre a Universidade do Minho e o Centro Hospitalar do Porto (CHP). Em constante evolução, o INTCare foi criado com o objetivo principal de ser um sistema inteligente capaz de prever a falha de órgãos, as suas consequências para os doentes e ainda sugerir tratamentos a realizar. Este objetivo levou à transformação do sistema de informação do hospital para um ambiente eletrónico e automático, já que grande parte da informação era inicialmente armazenada em formato de papel (Portela et al., 2014a).

Na sequência do projeto INTCare, surgiu este projeto de dissertação que está enquadrado na fase II do mesmo e tem como principal objetivo o desenvolvimento de uma *Pervasive Patient Timeline* interativa capaz de providenciar novo conhecimento útil, facilitando o processo de tomada de decisão em ambientes clínicos, sem restrições de local e tempo.

A introdução de características *pervasive* permite tirar partido de características como a ubiquidade, ou seja, torna-se possível aceder à informação a qualquer momento e em qualquer lugar; a capacidade de poder aceder à informação em tempo real; a automatização de procedimentos; a constante monitorização dos dados clínicos que promove o envio de alertas em caso de alterações negativas, entre outras. Também a incorporação de modelos *Data Mining* (DM) na *timeline* revela ser uma característica importante pois permite fazer previsões sobre situações futuras.

Como resultado, a *Pervasive Patient Timeline* apresenta-se como uma plataforma interativa, dinâmica e sempre disponível, podendo ser acedida em qualquer lugar e a qualquer momento,

que contém informações úteis (historiais clínicos e dados recolhidos em tempo real) para a decisão e a previsão de eventos clínicos.

A escolha deste tema deve-se ao gosto pela área de Sistemas Baseados em Conhecimento, em especial os sistemas de apoio à decisão que através do seu impacto nas mais diversas áreas de interesse são vistos como um complemento, não só à necessidade de tomar decisões, mas também às capacidades humanas, pelo que pretendem acima de tudo melhorar a qualidade de vida das pessoas.

Outro dos fatores de decisão é o interesse na área da medicina e na maneira como os sistemas de apoio à decisão podem melhorar a sua qualidade e desempenho. Entender o impacto que os sistemas de apoio à decisão podem ter nas UCI, como são explorados, de que forma facilitam as ações dos profissionais de saúde, que processos ajudam a simplificar, que tipo de decisões permitem justificar, entre outras.

Também a curiosidade ao nível do *output* final, a *Pervasive Patient Timeline*, revelou ser uma das razões da escolha deste tema. Primeiro por ser *pervasive*, na medida em que procura usar tecnologia que pode ser acedida em qualquer local, e através dos mais variados dispositivos, por forma a tornar as ações das pessoas mais simples, acessíveis e funcionais. E segundo, por a *timeline* ser uma abordagem única e inovadora, ao nível da saúde, que com recurso aos dados clínicos de um doente pretende apresentar a informação de uma forma diferente e interativa. Com o cruzamento de vários aspetos clínicos do doente, é assim permitido ao médico ter uma visão clínica diferente sobre o estado do doente, passando a ter uma base de decisão mais precisa.

## 1.2 Objetivos

O objetivo principal desta dissertação é desenvolver uma *pervasive timeline* interativa capaz de providenciar novo conhecimento de forma a facilitar os processos de tomada de decisão em ambientes clínicos, sem restrições de local e tempo. Desta forma, pretende-se responder à seguinte questão de investigação:

**De que forma as *pervasive timelines* podem contribuir para a melhoria da qualidade do processo de decisão, em Medicina Intensiva?**

De modo a ser possível atingir o objetivo principal, e por conseguinte dar resposta à questão de investigação, este objetivo foi desdobrado em objetivos gerais e específicos.

O objetivo geral é:

- Pesquisar e implementar o modelo de *Timeline* mais adequado ao apoio à decisão em Cuidados Intensivos;

Os objetivos específicos são:

- Desenhar modelo da *timeline* – definição de um conjunto de requisitos mínimos que as *timelines* a procurar devem possivelmente cumprir.
- Procurar e testar diferentes *timelines*;
- Escolher *timeline* entre as alternativas – analisar as *timelines* identificadas e perceber qual a melhor;
- Implementar *timeline* no sistema hospitalar – integração da *timeline* com o sistema de informação do hospital;
- Adicionar varáveis recolhidas em tempo real (sinais vitais, ventilação, ...) – permitir que seja possível visualizar dados de doentes do hospital na *timeline*;
- Adicionar o plano terapêutico – apresentar na *timeline* sugestões para determinado problema do doente;
- Adicionar previsões de *Data Mining* – passará pela adição de modelos de *Data Mining* à *timeline*, que sejam capazes de fazer previsões relativamente a algum aspeto futuro relacionado com o doente.

O desenvolvimento de um protótipo da *Pervasive Patient Timeline*, e sua implementação no atual Sistema INTCare permite uma maior automatização de procedimentos e criação de novo conhecimento, através da centralização de diversas fontes de informação e da apresentação da informação de uma forma mais perceptível e concisa, facilitando o exercício da atividade dos profissionais de saúde das Unidades de Cuidados Intensivos (UCI) do Hospital de Santo António.

### 1.3 Estrutura do Documento

Este relatório de dissertação encontra-se organizado, segundo a seguinte ordem de capítulos:

**Capítulo 1 – Enquadramento do Trabalho Proposto** – este capítulo apresenta uma breve introdução sobre as razões para o surgimento deste projeto e que necessidades vem suprimir. São também aqui reveladas as motivações da escolha deste trabalho, os objetivos do trabalho e ainda a estrutura do documento.

**Capítulo 2 – Revisão de Literatura** – este capítulo apresenta a estratégia de pesquisa utilizada para redigir a revisão de literatura, e aborda as temáticas e conceitos relacionados com o projeto a desenvolver. São abordados os seguintes temas, pela respetiva ordem: Estratégia de Revisão de Literatura, Medicina Intensiva e Unidades de Cuidados Intensivos, Sistemas de Apoio à Decisão, Descoberta e Conhecimento em Base de Dados, Data Mining, Sistema INTCare, Agentes Anti-Hipotensivos, *Pervasive Computing* e *A Timeline*.

**Capítulo 3 – Abordagem Metodológica** – nesta secção são explicitadas as metodologias utilizadas para o desenvolvimento do projeto. São apresentadas duas metodologias e um modelo: a Design Science Research (DSR), utilizada como metodologia de investigação que acompanha todo o trabalho. No desenvolvimento da solução, foi utilizado o Modelo do Processo de Tomada de Decisão desenvolvido por Herbert Simon, e mais recentemente adaptado por Turban, para o desenvolvimento de sistemas de apoio à decisão. Por último, foi usada a metodologia CRISP-DM para auxiliar no processo de Descoberta de Conhecimento em Base de Dados.

**Capítulo 4 – Desenvolvimento da Solução** – são explicitadas as tarefas executadas em cada uma das fases do modelo, que culminaram com o desenvolvimento da solução.

**Capítulo 5 – Descrição do Estudo de Data Mining** – neste capítulo é apresentado um estudo sobre a probabilidade de doentes da UCI terem de tomar vasopressores. Através da orientação da metodologia CRISP-DM foram desenvolvidos modelos de *Data Mining* com o propósito de previsão.

**Capítulo 6 – Discussão do Trabalho Realizado** – este capítulo tem como função discutir os resultados obtidos, de forma a verificar se estes cumprem os objetivos definidos.

**Capítulo 7 – Conclusão** – procura fazer um balanço do trabalho realizado e da forma como os seus constituintes contribuíram para que o resultado final fosse atingido. É ainda feita uma reflexão sobre os riscos que se verificaram ao longo do trabalho.

## **Capítulo 2 – Revisão de Literatura**

### **2.1 Estratégia de Revisão de Literatura**

De forma a perceber o contexto em que este projeto de dissertação se encontra inserido, foi necessário efetuar uma revisão de literatura. Através da revisão de literatura foram abordadas temáticas que se encontram relacionadas com o problema, com o intuito de ajudar a perceber o seu estado mais atual.

Para isso foi necessário efetuar uma pesquisa bibliográfica que abordasse os temas e áreas de estudo relacionados com o problema. Assim, foram reunidos alguns conceitos chave com o intuito de dar mais robustez à estratégia de pesquisa. Os conceitos são:

- *Intensive Medicine;*
- *Intensive Care Units;*
- *Decision Support Systems;*
- *Decision Support Systems in Intensive Care Units;*
- *Clinial Decision Support Systems*
- *Timeline;*
- *Pervasive Timeline in Intensive Medicine;*
- *Pervasive Computing;*
- INTCare.
- *Data Mining;*
- *Knowledge Discovery in Databases*
- *Vasopressors agents*
- *Inotropic agents*
- *Anti-Hipotensive Agents*

A pesquisa dos conceitos definidos foi realizada em motores de pesquisa *online* que devolvem, como resultado da pesquisa, publicações científicas. Os motores de pesquisa utilizados foram:

- Google Scholar;
- SpringerLink;
- ScienceDirect;
- b-on;
- RCAAAP;

- Repositório da Universidade do Minho.

Aquando da pesquisa, para uma revisão de literatura mais robusta e correta, foram tidos em conta como critérios de seleção das publicações a relevância do autor na área, o ano de publicação dos artigos e relevância dos artigos, tendo também sido realizadas pesquisas sobre os mesmos conceitos em outras línguas, tais como o português.

## **2.2 Medicina Intensiva e Unidades de Cuidados Intensivos**

Ao longo dos últimos 50 anos, foi possível observar o aparecimento e desenvolvimento de uma área das ciências médicas, que através de um esforço multidisciplinar nas áreas da fisiopatologia, terapêutica e tecnologias originou na Medicina Intensiva (Direção Geral de Saúde, 2003).

A Medicina Intensiva é uma área da medicina que se preocupa em diagnosticar e tratar doentes que sofrem de patologias que ameaçam as suas vidas, com o objetivo de recuperar o estado de saúde e qualidade de vida anteriores (Suter et al., 1994).

A necessidade de concentrar competências, conhecimento e tecnologias em locais capazes de cumprir os objetivos da Medicina Intensiva levou à criação de Unidades de Cuidados Intensivos (UCI) (Direção Geral de Saúde, 2003).

Os doentes em risco de vida encontram-se internados em UCI onde, através de vários equipamentos, se monitoriza constantemente o seu estado de saúde. Assim, a capacidade de salvar vidas, bem como a de alongar a vida daqueles que sofrem de patologias graves sem cura, aumenta (Silva, 2007). Enquanto a vida dos doentes se encontra ameaçada, estes são tratados recorrendo a medicação ou a dispositivos médicos, até recuperarem as funções vitais que os permitam viver de forma autónoma. Como exemplo, um doente com falha respiratória poder ser ligado a máquinas de ventilação, ou se tiver uma falha ao nível dos rins pode ser ligado a máquinas que substituam as funções dos rins, fazendo hemodiálise (Ramon et al., 2007).

Segundo Ramon (2007), é estimado que 70% dos doentes das UCI precisam de suporte de vida e monitorização apenas por uns dias, e têm uma taxa de sobrevivência alta. Enquanto 30% dos doentes ficam internados por longos períodos de tempo, por vezes meses. Estes doentes apresentam no geral um sistema imunitário fora de controlo, sujeitos a infeções graves e reações hiper-inflamatórias, sendo que à medida que o tempo de internamento aumenta a probabilidade de falecer também pode aumentar. No entanto, apesar das diversas descobertas que foram

surgindo ao nível dos sistemas de monitorização e tratamentos terem aumentado a taxa de sobrevivência nas UCI, ainda restam muitos desafios por ultrapassar.

A informatização das UCI, através dos avanços da tecnologia, forneceu os mais variados equipamentos capazes de estar constantemente a monitorizar diferentes funções vitais dos doentes e seus sistemas orgânicos, o que resulta no aumento da quantidade de dados disponíveis para melhor diagnosticar e tratar os doentes (Silva, 2007). Inicialmente, os equipamentos foram desenvolvidos para analisar sinais vitais como batimento cardíaco, pressão sanguínea, entre outros. Atualmente, nas UCI são medidos em tempo-real um elevado número de parâmetros associados à ventilação e sinais vitais (Geiger, 1985).

O aumento da quantidade de dados disponíveis nos cuidados intensivos leva a que exista um maior número de variáveis com que os profissionais de saúde têm de lidar para saberem qual a melhor decisão a tomar. Por vezes, podem ter que lidar com mais de duzentas variáveis (Morris and Gardner, 1992), o que dificulta a capacidade de conciliar e analisar tanta informação, por forma a decidir sobre qual a melhor ação a tomar.

É aqui que entram os sistemas de apoio à decisão, que tirando partido da elevada capacidade de processamento dos computadores procuram aliviar as dificuldades dos profissionais de saúde na interpretação da informação. Através da criação de novas informações resultante do cruzamento de imensas variáveis, os profissionais de saúde sentem-se assim mais capazes e seguros de tomar decisões mais precisas.

## **2.2 Sistemas de Apoio à Decisão**

Os Sistemas de Apoio à Decisão (SAD) são cada vez mais utilizados nos dias que correm, tendo expandido o seu nível de aplicação para diversas áreas de estudo. No início dos anos 70, começaram a ser articulados os primeiros conceitos relacionados com SAD, sob o termo de “Sistemas de Gestão de Decisão” (McKenney and Scott, 1984). Ainda que alguns autores considerem o termo SAD demasiado abrangente para ser facilmente definido, a definição mais comum na área de sistemas de informação é aquela que caracteriza os SAD como sistemas interativos baseados em computadores, que ajudam, quem tem de tomar decisões, a utilizar dados e modelos para resolver problemas não estruturados (Sprague and Ralph, 1980). Algumas das características observadas em SAD revelam que estes são direcionados para problemas menos estruturados e com poucas especificações; procuram combinar o uso de modelos ou técnicas analíticas com os dados tradicionais e funções de recolha; focam-se de uma maneira interativa

em aspetos que os tornam fáceis de usar por pessoas sem experiência com computadores e dão elevado ênfase à flexibilidade e adaptabilidade, com o intuito de acomodar alterações do ambiente e de abordagem na tomada de decisão do utilizador (Sprague and Ralph, 1980).

Ao desenvolvimento de um SAD, está normalmente associado um modelo de tomada de decisão que zela pela utilidade e eficiência do sistema desenvolvido.

O modelo mais popular é o inspirado pelo trabalho de Herbert Simon, e é constituído pelas fases Inteligência, Desenho, Escolha, Implementação (Simon, 1960) e Monitorização (Turban et al., 2014). Na fase Inteligência ocorre a análise, exploração e descrição do problema; na fase Desenho desenvolvem-se e analisam-se possíveis soluções para o problema; na fase Escolha procede-se à escolha da solução mais apropriada à resolução do problema; a fase Implementação, que embora não contemplasse no modelo de Simon como uma fase distinta, mas sim inserida na fase Escolha, foi defendida por vários autores como de relevo suficiente para ser uma fase independente, sendo que procura aplicar a solução escolhida ao problema em causa (Sprague and Ralph, 1980). Mais tarde foi adicionada por Turban (2014) a fase de Monitorização com a intenção de que fosse realizada uma avaliação sobre o processo de implementação, por forma a perceber se os resultados antecipados foram atingidos, efetuar as modificações necessárias e monitorizar a solução desenvolvida.

Em Medicina Intensiva, estes SAD são chamados de Sistemas de Apoio à Decisão Clínica (SADC). Os SADC são sistemas de computação desenhados para ter impacto na tomada de decisões clínicas sobre doentes, no exato momento em que as decisões precisam de ser tomadas (Berner and La Lande, 2007).

Alguns dos primeiros SADC a aparecer foram o Sistema de Dor Abdominal de Leeds (SDAL), o MYCIN e o *Health Evaluation through Logical Processes* (HELP). O SDAL usava informação sobre sensibilidade, especificidade e prevalência de doença para identificar vários sinais e sintomas, e resultados de testes para calcular sete possíveis razões para a dor abdominal aguda. O MYCIN centrou a sua atenção em problemas causados por infeções bacterianas, e atuava através de regras de produção, que através de determinadas condições apontavam para a possível infeção. O sistema HELP apresentava a capacidade de emitir alertas assim que eram detetadas anormalidades em relação à informação do doente, tendo servido como modelo de referência para o desenvolvimento de futuros SADC (Musen et al., 2006).

Os SAD já há algum tempo que fazem parte dos sistemas de informação das instituições hospitalares, no entanto sempre desempenharam, principalmente, funções de análise a

informação financeira e administrativa (Oliveira, 2001). Mas desde a publicação do Instituto de Medicina, *To Err is Human*, que têm sido apontados como elemento chave para aumentar a segurança dos doentes, através da minimização de erros médicos (Berner and La Lande, 2007). Ainda que o uso mais comum de SADC seja para suportar necessidades clínicas, têm também o potencial de diminuir custos, aumentar a eficiência e reduzir complicações médicas e administrativas para os doentes (Berner, 2009).

Para que um SADC seja bem-sucedido deve haver uma visão bem definida do propósito do sistema, de modo que este se adapte ao funcionamento do local onde vai ser implementado e às pessoas que o vão utilizar, e não o contrário. Segundo Khalifa (2014), algumas das estratégias a seguir para se ter um SADC implementado com sucesso devem ser:

- Incluir conteúdo correto, isto é, a transformação para formato digital de guias e protocolos clínicos do hospital, dos quais são derivadas regras a ser integradas com o sistema de informação hospitalar;
- Fornecer informação válida e relevante;
- Emitir mensagens simples, fáceis de responder e compreender, já que o sistema é mais eficiente quando diminui o esforço e tempo exercido pelos utilizadores para perceber as mensagens;
- Fornecer clarificação lógica que suporte a informação apresentada, aumentando assim a fiabilidade e credibilidade do sistema;
- A sua utilização deve servir como maneira de diminuir tempo e esforço, e não o contrário;
- Ser integrado com o funcionamento do trabalho clínico, caso contrário arrisca-se a não ter o efeito desejado e porventura deixa de ser usado;
- Adotar mecanismos de alerta passivos e ativos, especialmente ativos, já que informam automaticamente o utilizador sobre o que fazer sem estes terem que inquirir o sistema;
- Aumentar a capacidade de resposta e velocidade de funcionamento do sistema existente;
- Ser integrado com o sistema de informação do hospital, de modo a não haver necessidade de introduzir informação redundante;
- Existir um continuado esforço de manutenção relacionado com o funcionamento do SADC;

Os SADC podem se basear em regras de produção ou utilizar *machine learning* e inteligência artificial (Berner and La Lande, 2007). Da tentativa de construir um programa de computador capaz de simular a maneira de pensar humana, surgiram os sistemas baseados em conhecimento

(Shortliffe et al., 1973). Neste tipo, o sistema encontra-se dividido em três partes: a base de conhecimento, o motor de inferência e um mecanismo de comunicação com o utilizador. A base de conhecimento consiste em informação compilada, normalmente, na forma de regras *if-then*. O motor de inferência contém as fórmulas para combinar as regras da base de conhecimento com a informação do doente, e o mecanismo de comunicação encarrega-se de apresentar os resultados ao utilizador (Berner and La Lande, 2007).

Por sua vez, os sistemas não-baseados em conhecimento usam uma forma de inteligência artificial denominada *machine learning*, que permite o computador aprender com experiências passadas e reconhecer padrões nos dados (Marakas, 2003).

Contudo, e ainda que os benefícios da implementação de SADC sejam muitos, também existem fatores que atuam como barreiras à sua implementação e podem fazer com que não seja tirado o maior proveito de cada um deles, tornando assim a implementação do SADC pouco vantajosa, já que dificilmente será tirado real valor da sua utilização. Para além de barreiras financeiras, estas podem também ser categorizadas como humanas, técnicas e organizacionais. Segundo (Trivedi et al., 2002), alguns dos fatores que representam estas categorias e podem impedir a implementação de SADC com sucesso são:

- A perceção que os profissionais de saúde têm da utilidade de computadores em ambiente clínico;
- O conhecimento e atitude dos profissionais de saúde em relação a computadores;
- O conhecimento dos profissionais de saúde sobre o sistema;
- A facilidade de uso;
- A utilidade e flexibilidade do sistema;
- A aplicabilidade e adaptabilidade do sistema ao ambiente de trabalho;
- A qualidade do treino para aprender a usar o programa e da informação fornecida;
- O envolvimento dos profissionais de saúde no desenvolvimento e implementação;
- O envolvimento e apoio de profissionais de saúde mais experientes no processo de implementação;
- O clima organizacional do ambiente de trabalho;
- A existência de problemas técnicos e a qualidade do suporte fornecido para os resolver.

Como exemplo de um caso que falhou devido à existência de barreiras na implementação foi o do programa HealthSMART. Este programa foi implementado no estado de Victória na Austrália e

visava criar um sistema eletrónico de registos de saúde, mas foi cancelado em 2012 devido a problemas financeiros (Charette, 2012).

Por outro lado, no estado da Austrália do Sul decorre um processo de implementação do *Enterprise Patient Administration System* (EPAS), que irá permitir que os registos de informação de saúde sejam totalmente eletrónicos. A informação disponibilizada será relativa ao plano de tratamento, a resultados de exames e a medicação que o doente esteja a tomar. Tem também a capacidade de oferecer acesso em tempo real à informação sobre o doente na altura em que as decisões têm de ser tomadas. Para a projeção e implementação do sistema foram consultados mais de 350 funcionários clínicos e administrativos, e mais de 100 estiveram envolvidos na fase de testes do sistema, assegurando que este atingisse as necessidades do sistema de saúde pública (SA Health). Durante 2013 e inícios de 2014, o EPAS já foi implementado em 7 hospitais e centros de saúde, contando com mais de 5000 funcionários que usam o EPAS diariamente. Os benefícios clínicos que advém da implementação do EPAS são, nomeadamente, a emissão de alertas relacionados com a dosagem de medicação, alergias e interações entre medicamentos; melhoria nos cuidados dos doentes, devido ao acesso mais rápido à informação do doente e à possibilidade de mais do que um utilizador poder aceder à informação do mesmo doente, o que não seria possível se esta estivesse em formato de papel; acesso à informação do doente por parte do serviço de ambulâncias e envio de alertas, através de mensagem de telemóvel, para os doentes sobre a aproximação de uma consulta (SA Health).

O desenvolvimento deste projeto está associado a um SADC, o sistema INTCare. A criação deste projeto deve-se à dificuldade existente dos profissionais de saúde (médicos e enfermeiros) em lidar com informação proveniente de monitores de sinais vitais, ventiladores, farmácia, laboratórios clínicos, folha de enfermagem eletrónica, entre outros. A informação resultante destas fontes pode vir representada de diversas formas, sendo que algumas delas são gráficos, tabelas e texto. A apresentação da informação depende da forma como cada um desses sistemas irá expor os resultados do seu processo. Se a informação for proveniente do laboratório, serão recebidos exames médicos com o seu tipo de apresentação padrão associado, se for de sistemas de monitorização do doente, a forma de apresentar a informação já será outra. O mesmo raciocínio se aplica às outras fontes que fornecem informação/dados.

Tendo em vista uma solução para este problema de heterogeneidade existente entre diferentes dados clínicos, foi fundada em 1987, a *Health Level 7 International*. É uma organização internacional de desenvolvimento de *standards* que possibilita a interoperabilidade da informação

clínica, através do estabelecimento do *Health Level 7* (HL7), uma linguagem padronizada para a partilha, gestão e recolha de dados clínicos, que irão servir de suporte à atividade clínica e à gestão, entrega e avaliação dos serviços de saúde (Health Level Seven International, 2007).

Visto que os profissionais de saúde precisam de lidar com muita informação proveniente de diferentes locais ao mesmo tempo, existe a necessidade de suprimir a diversidade entre a informação, que obriga os profissionais de saúde a terem de interagir com diversos sistemas, até conseguirem perceber o estado atual do doente. A dificuldade no cruzamento de um elevado número de informação com o objetivo de deliberar sobre uma decisão a tomar também revela ser uma necessidade que precisa de ser satisfeita.

Assim, é da constatação destas necessidades que foi criado este projeto, que pretende através da criação de uma plataforma capaz de interoperar com os diversos sistemas hospitalares e apresentar toda a informação recolhida e fundamental para o processo de decisão, independentemente da sua fonte de dados. Para isso, será integrada uma *pervasive timeline*, no atual sistema INTCare, que irá receber a informação clínica de diferentes locais e irá apresentá-la no interface da *timeline*, desde o momento do internamento do doente até ao momento em que receber alta. Desta forma, os profissionais de saúde podem visualizar toda a informação num só local, podendo analisar e deliberar uma decisão mais rapidamente, com maior eficiência e eficácia. Assim, e tendo em conta todas estas alterações inerentes à implementação de SADC, percebe-se que os benefícios futuros podem ser muitos, pelo que deve-se trabalhar em tentar diminuir o impacto das forças opostas à sua implementação. Desta forma e com base na revisão de literatura efetuada é possível afirmar que os SADC podem alterar a forma como se ensina e pratica medicina.

### **2.3 Descoberta de Conhecimento em Base de Dados**

A Descoberta de Conhecimento em Base de Dados (DCBD) surgiu devido ao aumento exponencial da quantidade de dados, bem como a acessibilidade aos mesmos, sendo a facilidade do acesso eletrónico aos dados um dos principais responsáveis. Desde a era da informação que a acumulação de dados é cada vez mais fácil e o seu armazenamento barato, pelo que estima-se que a quantidade armazenada duplique a cada vinte meses (Paiva et al., 2012). Contudo, a capacidade de perceber a utilidade desses dados não consegue acompanhar o ritmo de crescimento verificado da quantidade de dados disponíveis eletronicamente, verificando-se cada vez mais uma maior diferença entre a capacidade de perceção e o número de dados disponíveis.

Desta forma é necessário recorrer ao processo de DCBD, para encurtar a distância entre a nossa incapacidade de conseguirmos aumentar o nosso nível de processamento e a elevada quantidade de dados disponíveis.

A DCBD é um processo automático de exploração que analisa e faz modelação de repositórios com elevada quantidade de dados. É um processo organizado que procura identificar padrões válidos, novos, úteis e compreensíveis em conjuntos de dados com complexidade e quantidade elevada.

O processo de DCBD é iterativo na medida em que há a possibilidade de retroceder no processo por forma a melhorar as etapas previamente executadas, e é interativo pois a realização de atividades tem efeitos diretos sobre os dados em análise.

O processo começa pelo entendimento do domínio em análise e pela definição dos objetivos a atingir com a implementação do processo, dando de seguida início às etapas onde serão efetuadas transformações aos dados. Este processo inspirado em Fayyad et al. (1996) e aqui descrito por Maimon and Rokach (2005) consiste em nove etapas como se pode ver na Figura 1:

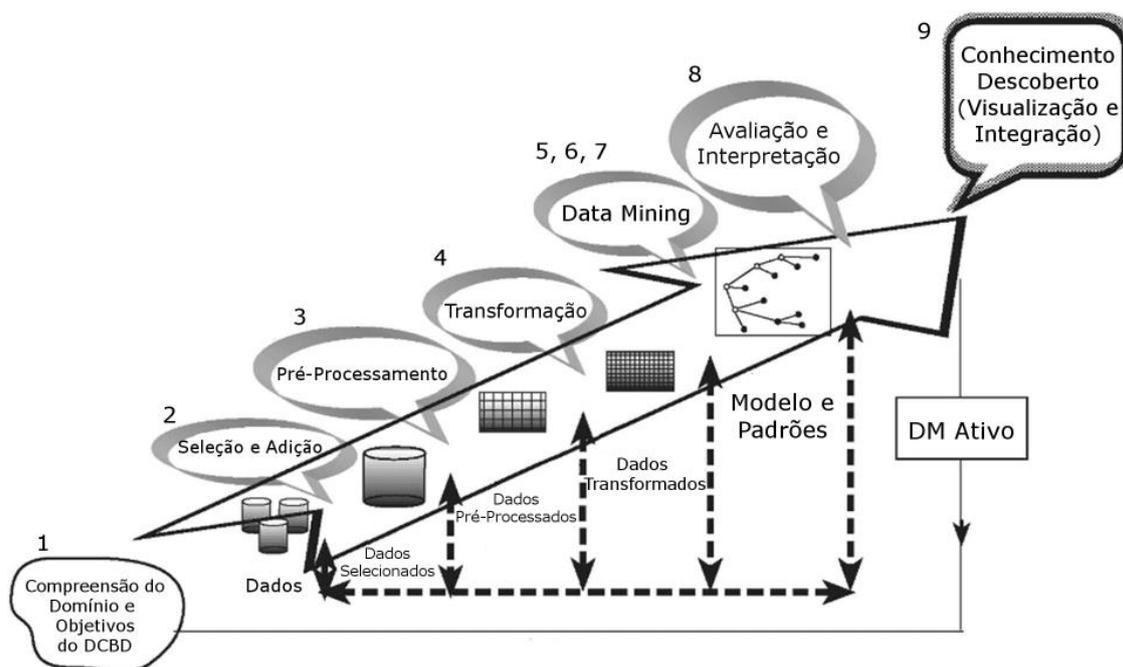


Figura 1 – Processo de DCBD (adaptado de Maimon and Rokach, 2005)

### 1. Desenvolver um entendimento do domínio de aplicação

É uma etapa preparatória de todo o processo que se foca essencialmente em perceber qual o domínio e contexto em que o projeto/trabalho se irá desenrolar. É também nesta etapa que

são definidos os objetivos, bem como o ambiente em que o processo, assim que finalizado, será aplicado.

## **2. Selecionar e criar um conjunto de dados onde será feita a descoberta**

Após a definição dos objetivos, deve-se determinar que dados vão ser usados para o processo de descoberta de conhecimento. Este processo envolve alguns aspetos, nomeadamente, perceber que dados estão disponíveis, se há necessidade de mais dados, e se é possível integrá-los todos no mesmo conjunto de dados. Esta etapa é de elevada importância porque a qualidade dos resultados está diretamente relacionada com a qualidade dos dados usados, de modo a que deve-se considerar o maior número de atributos, pois caso falem atributos relevantes ao processo, o resultado final pode não ser o esperado. Apesar dos custos associados à recolha, organização e exploração de repositórios de dados complexos há uma compensação na oportunidade de melhor perceber como os dados podem afetar o processo.

## **3. Pré-processamento e limpeza**

É aqui onde se procura aumentar a qualidade e fiabilidade dos dados. Para isto efetuam-se ações que visam limpar os dados de incorreções, nomeadamente, lidar com valores em falta e remoção de *outliers*.

## **4. Transformação dos dados**

Nesta etapa, os dados são processados e preparados para serem utilizados pelos modelos de *Data Mining* (DM). Os métodos aqui utilizados envolvem a redução da dimensão (*feature selection*, extração e *record sampling*), e transformação do atributo (discretização de atributos numéricos e transformação funcional). Esta etapa pode ser crucial para o sucesso do processo e é normalmente muito específica, ou seja, é afetada pelo ambiente em que se pretende aplicar o processo de DCBD.

## **5. Escolher o tipo de modelo *Data Mining* apropriado**

A partir desta etapa dá-se início ao processo de indução de modelos de DM. Aqui se escolhe qual o modelo mais apropriado a executar, se classificação, regressão, *clustering*, entre outros. Esta decisão depende dos objetivos definidos para o processo de DCBD e das decisões tomadas nas etapas anteriores. O objetivo definido para a realização de DM pode ser de dois tipos: a previsão de dados ou descrição de dados.

## **6. Escolher o algoritmo de *Data Mining***

Assim que for escolhida o tipo de tarefa a utilizar, procede-se à seleção do algoritmo de *Data Mining* mais apropriado. A existência de vários algoritmos leva à necessidade de perceber em

que aspetos é que uns são melhores que outros, de modo a que se consiga obter os melhores resultados possíveis.

### **7. Aplicar o algoritmo de *Data Mining***

É nesta etapa que o algoritmo escolhido é executado. Contudo este pode ter de ser executado várias vezes até se conseguir encontrar os parâmetros em que o algoritmo apresenta melhores resultados.

### **8. Avaliar resultados**

Após a obtenção dos resultados, deve ser feita uma avaliação dos mesmos tendo em conta os objetivos definidos. Caso os resultados não sejam os esperados deve-se reavaliar as etapas anteriores e reiterar, caso se justifique, para que possa haver um aumento na qualidade dos resultados. É também efetuada a documentação do conhecimento descoberto com o intuito de este ser usado no futuro caso seja necessário.

### **9. Usar a descoberta de conhecimento obtida**

Esta é a última etapa do processo e tem por base a incorporação do conhecimento descoberto no ambiente para o qual se destinava. O conhecimento torna-se ativo na medida em que o podemos ver em ação e medir os seus efeitos. O nível de sucesso desta etapa determina o nível de eficácia de todo o processo previamente executado. É também uma etapa de vários desafios, que devido à aplicação de novo conhecimento pode levar a que sejam efetuadas alterações na forma de operar, que até então não eram consideradas.

A aplicação deste processo traz claros benefícios já que nos permite obter significados para dados que até ao momento não eram considerados. Este processo tem aplicação nas mais diversas áreas sendo que as mais comumente associadas são o marketing, deteção de fraudes, finanças, investimentos, telecomunicações e negócios (Fayyad et al., 1996). No que toca à medicina intensiva também se tem verificado um aumento da aplicação do processo de DCBD de forma a reduzir a incerteza na tomada de decisões em ambiente clínico (Meyfroidt et al., 2009).

## **2.4 *Data Mining***

O *Data Mining* (DM) apresenta como um dos principais objetivos a conversão de dados em informação/conhecimento útil. Mais especificamente, o DM é um processo de procura de padrões e relações em grandes quantidades de dados, tendo como propósito a descrição ou previsão. A descrição foca-se na procura de padrões, e posterior apresentação dos mesmos de uma forma

compreensível. Enquanto a previsão envolve a identificação de variáveis na base de dados que possam ser úteis para prever acontecimentos futuros (Hardin and Chhieng, 2007). No entanto, as fronteiras nem sempre são bem definidas, isto porque, modelos de previsão podem ser descritivos na medida em que estes são objeto de compreensão, e vice-versa. Estes objetivos de descrição ou previsão podem ser atingidos recorrendo a uma variedade de modelos de DM (Fayyad et al., 1996).

Os modelos de DM podem ser classificados essencialmente em seis tipos: Regressão, Classificação, *Clustering*, Sumarização, Modelação em Dependência, Detecção de Desvios e Alterações (Fayyad et al., 1996). A Regressão procura mapear um item do conjunto de dados a uma variável de previsão com um valor real; a Classificação é um modelo que se destina a classificar uma variável do conjunto de dados, tendo em conta um grupo predefinido de classes; o *Clustering* é um modelo que se foca em identificar um conjunto finito de categorias para descrever os dados de um conjunto; a Sumarização, também ela uma tarefa descritiva, envolve métodos que pretendem fazer uma representação compacta de um conjunto de dados; a Modelação em Dependência foca-se em encontrar um modelo que seja capaz de descrever dependências significantes entre as variáveis de um conjunto de dados e a Detecção de Desvios e Alterações dedica-se a encontrar as mudanças mais significativas nos dados de valores previamente medidos ou normativos.

A estes modelos são aplicadas técnicas de DM que irão atuar sobre o conjunto de dados, analisando-o e apresentando posteriormente os resultados relativos aos modelos desenvolvidos, e que servirão para avaliar a qualidade dos modelos. A existência de diversas técnicas permite realizar várias tentativas de forma a perceber qual a técnica que se aplica melhor ao conjunto de dados, no entanto algumas técnicas podem à partida ser descartadas pois a sua aplicação depende do tipo de objetivo do DM, isto é, previsão ou descrição.

Dado que os objetivos do DM que se perspetivam neste trabalho são a previsão, e por conseguinte a aplicação de modelos de classificação apresenta-se algumas das principais técnicas aplicadas nestes modelos:

As Árvores de Decisão representam um conjunto de regras estruturadas através de uma hierarquia de classes ou valores. O nome advém da sua estrutura ser constituída por nós, onde se especifica um teste para os atributos da variável, e ramos, que correspondem a um valor possível para o atributo em causa (Santos and Azevedo, 2005).

A Naïve Bayes é uma técnica que tem a capacidade de prever probabilidades de um atributo pertencer a uma determinada classe. Para isso é usado um classificador denominado Naïve Bayes que assume que o efeito do valor de um atributo em uma determinada classe é independente dos valores dos outros atributos. A origem desta técnica advém do teorema de Bayes que de uma forma sucinta descreve a probabilidade de um evento acontecer baseando-se nas condições relacionadas com esse evento (Han et al., 2011). O teorema é definido através da seguinte equação (Bayes and Price, 1763):

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}, \quad \text{onde A e B são eventos}$$

- $P(A)$  e  $P(B)$  são as probabilidades de A e B acontecerem sem se influenciarem;
- $P(A|B)$ , sendo probabilidade condicional, é a probabilidade de A sendo que B é verdade;
- $P(B|A)$ , é a probabilidade de B sendo que A é verdade.

*Support Vector Machines* (SVM) é um método usado na classificação de dados lineares ou não-lineares. Faz uso de um mapeamento não linear para transformar os dados de treino originais para uma dimensão superior. Nesta nova dimensão o SVM procura pelo hiper-plano de separação ótimo, recorrendo a vetores de suporte e a margens definidas por esses vetores. Desta forma os dados com dois tipos de classes diferentes podem ser separados e classificados (Han et al., 2011). As Redes Neurais Artificiais são modelos altamente simplificados da estrutura da rede neuronal biológica. Consistem na interconexão de unidades de processamento que de acordo com uma topologia tentam realizar tarefas de reconhecimento de padrões. Estas unidades de processamento recebem valores como *input*, avaliam-no e retornam uma soma desses valores, sendo de seguida emitido um sinal resultante da soma (Yegnanarayana, 2009).

Os Modelos Lineares Generalizados são de forma sintética a aplicação da regressão linear em modelos que se focam em prever categoricamente a resposta das variáveis. Neste caso, a variância da variável de resposta  $\alpha$  é uma função do valor médio de  $\alpha$ , enquanto na regressão linear o valor de  $\alpha$  é constante. (Santos and Azevedo, 2005).

De forma a obter boas estimativas nos modelos são usados métodos durante a sua execução com o intuito de os testar e treinar. Um dos métodos mais comuns é o *Cross Validation* que divide os dados em  $k$  segmentos. A cada vez, um dos segmentos  $k$  é usado para o conjunto de teste e os restantes  $k-1$  segmentos são usados para o conjunto de treino. Assim cada dado tem a oportunidade de estar no conjunto de teste uma vez e  $k-1$  vezes no conjunto de treino. Desta forma

a distribuição treino/teste é executada  $k$  vezes. A variância da probabilidade estimada é assim reduzida à medida que  $k$  aumenta, mas por outro lado o tempo de execução aumenta  $k$  vezes (Schneider, 1997).

Depois de executados os modelos, vem de seguida a fase de avaliação dos resultados que é das fases mais importantes do processo de DM, pois é através da avaliação que se consegue identificar o modelo que apresenta melhores resultados. Nesta fase é necessário promover o uso de métricas de avaliação adequadas já que estas têm influência direta na perceção que temos dos resultados. Em modelos de classificação, as métricas mais utilizadas são a Matriz de Confusão e as curvas designadas *Receiver Operating Characteristic* (ROC). A matriz de confusão é uma tabela com duas colunas e duas linhas onde são representados o número de classificações reais *versus* o número de classificações previstas para cada uma das classes do conjunto de dados. De forma a definir as possibilidades de acerto e erro para as classes são usados quatro indicadores:

- **Verdadeiros positivos** (VP) – valores positivos classificados como tal;
- **Verdadeiros negativos** (VN) – valores negativos classificados como tal;
- **Falsos positivos** (FP) – valores negativos classificados como positivos;
- **Falsos negativos** (FN) – valores positivos classificados como negativos.

Na Tabela 1 é possível ver representada uma matriz de confusão.

**Tabela 1 – Matriz de Confusão**

		Previsão	
		Positivos	Negativos
Real	Positivos	Verdadeiros Positivos	Falsos Positivos
	Negativos	Falsos Negativos	Verdadeiros Negativos

A partir destes indicadores da matriz de confusão é possível definir um conjunto de métricas que servem para aferir a performance classificativa dos modelos. Três das principais métricas são a sensibilidade, especificidade e acuidade.

A sensibilidade mede a proporção de valores verdadeiros positivos e é dada pela expressão:

$$Sensibilidade = \frac{VP}{(VP + FN)}$$

A especificidade mede a proporção de valores verdadeiros negativos e é dada pela expressão:

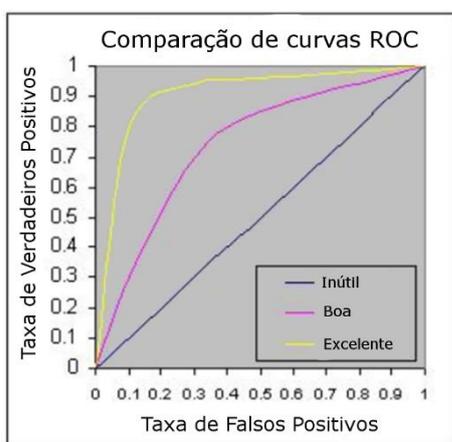
$$Especificidade = \frac{VN}{(VN + FP)}$$

A acuidade mede a proporção de valores verdadeiros positivos e negativos e é dada pela expressão:

$$Acuidade = \frac{VP + VN}{(VP + VN + FN + FP)}$$

Por outro lado temos as curvas ROC que são ilustrações gráficas da performance de sistemas de classificação binária. A criação da curva é dada pela relação existente entre a taxa de verdadeiros positivos e a taxa de falsos positivos. As análises à curva ROC permitem também observar a relação existente entre a sensibilidade e especificidade, isto é, um aumento na sensibilidade resulta num decréscimo da especificidade. A avaliação à curva é feita através da área sob a curva, sendo que quanto maior for a área por baixo da curva mais preciso é o modelo ao qual a curva pertence, significando isto que à medida que a área aumenta, também a capacidade do modelo em separar os valores de uma classe da outra aumenta. Na Figura 2 pode-se observar a representação gráfica de três curvas ROC e na Tabela 2 pode-se ver a relação existente entre os valores da área sob a curva e capacidade classificativa do modelo (Tape, 2006).

**Tabela 2 – Relação entre área e capacidade classificativa**



Área	Capacidade Classificativa
[0,5;0,6[	Inexistente
[0,6;0,7[	Fraca
[0,7;0,8[	Razoável
[0,8;0,9[	Boa
[0,9;1]	Excelente

**Figura 2 – Exemplo de curvas ROC (adaptado de Tape, 2006)**

O uso destas métricas de avaliação nos modelos irá permitir tirar conclusões relativamente à qualidade dos modelos criados, à forma como os dados usados estão a afetar os resultados obtidos, que dados tem mais preponderância, quais os melhores modelos por comparação entre eles, entre outras observações.

O uso das técnicas de DM tem-se tornado cada vez mais presente na área da medicina devido ao impacto que podem ter na melhoria da qualidade dos serviços prestados. A presença de grandes quantidades de dados é também um dos fatores que torna a medicina como uma área propícia a oportunidades de implementação de modelos de DM. Há estudos que usam DM para prever quando e como alterar a dosagem de um medicamento (Morik et al., 1999), reinternamentos (Braga et al., 2014), o *outcome* da condição de um doente internado (Santos et al., 2011a), a falência de órgãos (Santos et al., 2010), a duração da estadia (Velooso et al., 2014), receção de alta hospitalar (Portela et al., 2014b), a sobrevivência ao cancro da mama (Delen et al., 2005), doenças cardíacas (Soni et al., 2011), a sobrevivência a operação de transplante de coração e pulmão (Oztekin et al., 2009), entre muitas outras aplicações que vão aparecendo com a finalidade de melhorar a prática da medicina.

## 2.5 Sistema INTCare

O INTCare é um projeto de investigação, que resulta da parceria conjunta entre a Universidade do Minho e o Centro Hospitalar do Porto (CHP), e encontra-se implementado na Unidade de Cuidados Intensivos (UCI) do Hospital de Santo António, no Porto.

Tendo começado no ano de 2008, uma das principais preocupações foi transformar o processo de aquisição de dados do sistema de informação do hospital. Até então, os registos eram efetuados em papel e manualmente, mas com o INTCare os processos evoluíram para um sistema de recolha automático e eletrónico (Portela et al., 2011).

O INTCare é um Sistema *Pervasive* de Apoio à Decisão (SPAD) que faz uso de agentes inteligentes para processar as suas tarefas. Atuando de forma automática e em tempo real, tem como finalidade ajudar os profissionais de saúde a decidir o melhor tratamento a dar aos doentes, procurando oferecer suporte a decisões clínicas, na medida em que faz previsões da falência de órgãos e suas consequências, e sugere procedimentos, tratamentos baseados nas previsões realizadas; tudo isto através de técnicas de *Data Mining* (DM) ((Santos et al., 2011b) (Portela et al., 2014a)).

O sistema encontra-se dividido em quatro subsistemas (Figura 3): Aquisição de Dados, Gestão de Conhecimento, Inferência e Interface, constituídos por agentes inteligentes que cooperam entre si (Santos et al., 2011b).

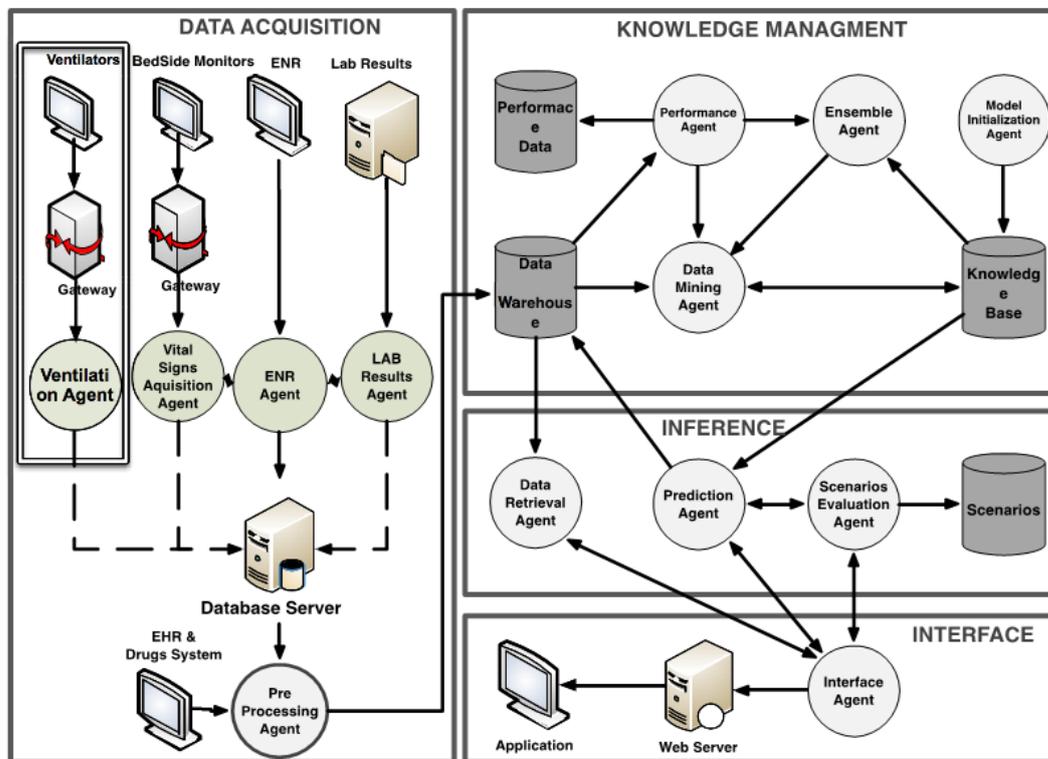


Figura 3 – Sistema INTCare e subsistemas (retirado de Oliveira et al., 2014)

As atividades de aquisição de dados são executadas no subsistema Aquisição de Dados, que redireciona a informação para um *Data Warehouse*, em um formato apropriado para os agentes do subsistema Gestão de Conhecimento. Por sua vez, o subsistema Gestão de Conhecimento mantém os modelos de previsão usados pelo subsistema Inferência, não descurando a sua validade e atualizando-os sempre que indicado. Por fim, o subsistema Interface é onde se observa a interação dos profissionais de saúde com o sistema (Santos et al., 2006).

Os dados que o sistema INTCare processa são provenientes de cinco fontes: Monitores de Sinais Vitais, Ventiladores, Folha de Enfermagem Eletrônica, Processo Clínico Eletrônico, Laboratórios e Farmácia (Portela et al., 2014a).

O sistema apresenta a capacidade de identificar eventos críticos relacionados com os doentes. A monitorização é automática e em tempo real, sendo exercida sobre as variáveis: pressão sanguínea, ritmo cardíaco, temperatura, produção de urina e saturação de oxigénio. O sistema emite um alerta sempre que estas variáveis saem do intervalo de valores normais por um período de tempo

prolongado ou sempre que se afastam demasiado do intervalo de valores normais. Os resultados são posteriormente apresentados em tabela e gráfico. Paralelamente é utilizado um sistema de cores para alertar sobre a condição do doente quando as variáveis estão fora do intervalo de valores normais. Esta informação pode ser acedida pelos profissionais de saúde, em qualquer lugar e a qualquer momento (Portela et al., 2014a).

Através da indução automática de modelos de DM tornou-se possível a realização de previsões sobre o estado dos doentes para a próxima hora, tudo isto em tempo real. Os modelos são capazes de se adaptar e otimizar ao longo do tempo, certificando assim que, sempre que requisitadas, as previsões apresentadas são as melhores possíveis (Portela et al., 2014a).

Este conjunto de transformações efetuadas no sistema de informação possibilitaram a criação de novo conhecimento útil para o processo de tomada de decisão, tornando o funcionamento das UCI mais eficiente e eficaz, diminuindo assim o grau de incerteza das decisões tomadas pelos profissionais de saúde e *stress* associado.

A realização deste projeto encontra-se no âmbito da fase II do INTCare que tem como principal objetivo tornar mais eficiente e eficaz o processo de tomada de decisão através da criação de novo conhecimento, sendo, neste caso, conseguido com o desenvolvimento de uma nova funcionalidade: a *Pervasive Patient Timeline*.

## 2.6 Agentes Anti-Hipotensivos

O uso de agentes inotrópicos e vasopressores tem-se vindo a tornar um pilar na gestão de importantes síndromes vasculares. No geral estes agentes têm efeitos excitatórios e inibidores ao nível do coração e músculos vasculares, bem como importantes efeitos metabólicos, no sistema nervoso central e no sistema nervoso autónomo pré-sináptico. Os inotrópicos positivos têm este efeito pois aumentam a contractilidade cardíaca, o que aumenta o volume de sangue ejetado e por conseguinte a pressão arterial. No caso dos vasopressores há uma contração das artérias, o que significa menor espaço para a mesma quantidade de sangue, resultando num aumento da pressão arterial (Overgaard and Džavik, 2008).

A este nível temos o grupo das catecolaminas que mediam as ações cardiovasculares através de recetores  $\alpha$ ,  $\beta$  e dopaminérgicos, os quais conforme a densidade e proporção modulam as respostas fisiológicas dos inotrópicos e vasopressores em tecidos individuais. Como parte deste grupo temos a Dopamina, Dobutamina, Noradrenalina, Adrenalina, entre outros (Overgaard and Džavik, 2008).

A ação destes agentes é exercida quando estes se ligam a receptores adrenérgicos  $\beta$  ou  $\alpha$  das células, estimulando o sistema nervoso simpático.

Este projeto irá concentrar-se em três destas substâncias simpatomiméticas, isto é, drogas que imitam os efeitos das substâncias de neurotransmissores do sistema nervoso simpático, e na sua capacidade de aumentar a pressão arterial. Estas drogas são adrenalina, noradrenalina e dopamina.

A adrenalina é uma catecolamina endógena com grande afinidade por receptores  $\beta$  e  $\alpha$ . Os efeitos ao nível dos receptores  $\beta$  são mais pronunciados em doses mais baixas, enquanto nos  $\alpha$  são precisas doses mais elevadas. A adrenalina é normalmente usada para tratar a paragem cardíaca, asma e anafilaxia (Medscape, 2015b).

A dopamina é um neurotransmissor endógeno central que atua nos receptores adrenérgicos e dopaminérgicos e provoca variados efeitos clínicos. Dependendo das doses a dopamina pode ter diferentes efeitos, a doses médias normalmente aumenta a contratilidade cardíaca e por conseguinte o ritmo cardíaco. Em doses mais elevadas a dopamina atua nos receptores adrenérgicos  $\alpha_1$  e provoca a vasoconstrição (Overgaard and Džavík, 2008). Esta é assim muitas vezes utilizado para tratar a hipotensão aguda, que é um resultado da pressão arterial baixa (Medscape, 2015a).

Noradrenalina é um neurotransmissor endógeno libertado pelos nervos adrenérgicos pós-ganglionares. É um potente agonista de receptores adrenérgicos  $\alpha_1$  e um modesto agonista de receptores  $\beta$ , o que o torna em um potente vasoconstritor com fracas capacidades inotrópicas positivas, ou seja, aumenta a pressão arterial através da constrição das artérias tendo um impacto mínimo ao nível do ritmo cardíaco. Isto torna a noradrenalina atrativa em situações que o aumento do batimento cardíaco seja indesejado (Overgaard and Džavík, 2008). Normalmente a noradrenalina é usada para tratar a sépsis, paragem cardíaca e choque séptico (Medscape, 2015c).

A utilização de DM neste projeto tem como principal objetivo a perceção da capacidade destes agentes conseguirem elevar a pressão arterial, pois pretende-se prever, através de determinados indicadores, se haverá necessidade de administrar um destes agentes.

## **2.7 Pervasive Computing**

Mark Weiser (1991) uma vez disse “As tecnologias mais profundas são aquelas que desaparecem. Elas tecem-se no tecido da vida quotidiana até que se tornem indistinguíveis.”. Isto significa que

as tecnologias que estiverem tão bem enraizadas com o dia-a-dia se tornam tão indispensáveis que deixam de ser vistas como uma entidade separada, ou seja, passam a incorporar o estilo de vida das pessoas.

Esta visão de Mark Weiser representa os fundamentos do *pervasive computing*. A ubiquidade é parte integrante do *pervasive computing*, sendo que permite o acesso em qualquer lugar e a qualquer momento, quer seja através de dispositivos fixos ou móveis, enquanto o *pervasive* é algo que significa estar presente no ambiente de forma impercetível (Varshney, 2009).

O *pervasive computing* tem assim como objetivo criar ambientes inteligentes onde os dispositivos que estão interligados e inseridos no ambiente oferecem uma conectividade contínua, fiável e não intrusiva, e serviços de valor acrescentado. O resultado é melhoria da experiência humana e qualidade de vida, sem uma percepção explícita da subjacente comunicação entre a tecnologia (Cook and Das, 2007).

Esse ramo está intimamente relacionado com os ambientes inteligentes, nos quais a tecnologia aplica inteligência artificial e técnicas de *machine learning*, de forma a conseguir aumentar a capacidade de deliberar, controlar e adaptar o meio (Lewis, 2004). Também nos sistemas ciberfísicos, que englobam sistemas centrados em computadores e informação como integração de comunicação, computação e controlo, pode haver exploração destas tecnologias, ainda que fora do contexto da experiência humana (Cook and Das, 2012).

É na noção da utilização de tecnologia que fornece capacidades de computação e comunicação de tal forma impercetíveis que os utilizadores deixam de se aperceber da sua utilização, que se distingue o *pervasive computing* da computação móvel e distribuída (Cook and Das, 2012).

Estas características do *pervasive computing* possibilitam a implementação de tecnologias *pervasive*, de quatro formas diferentes. A implantação de dispositivos no corpo humano; dispositivos associados a roupa, acessórios, entre outros; dispositivos portáteis e dispositivos ambientais (Varshney, 2009). O nível de impercetibilidade varia nestes quatro tipos, sendo que os portáteis poderão ser os mais visíveis, enquanto dispositivos associados a roupa e acessórios poderão ser pouco visíveis e os dispositivos implantados poderão ser invisíveis; já nos dispositivos ambientais o nível de impercetibilidade poderá variar bastante dependendo da sua aplicação (Varshney, 2009).

Ao longo das duas últimas décadas as tecnologias têm evoluído em direção a infraestruturas mais ubíquas e *pervasive*. Reconhecido como exemplo é o caso dos dispositivos móveis, como *smartphones*, *tablets*, entre outros, que recolhem, armazenam e divulgam informação, e

apresentam cada vez mais diversidade na sua aparência, portabilidade, recursos que oferecem e uso. Em adição, a sua capacidade de comunicação tem-se tornado cada vez mais rápida, robusta e *pervasive*. A visão do *pervasive computing* é tão poderosa que pretende atingir grandes quantidades de pessoas, como nunca a tecnologia conseguiu no passado (Cook and Das, 2012). Parte dessa visão é a criação de sistemas *pervasive* de larga escala, sendo aqui a escalabilidade o componente principal. No entanto, à medida que os sistemas crescem aumenta a heterogeneidade existente entre os dispositivos, aumentando assim o nível de complexidade de gestão do sistema (Cook and Das, 2012). Alguns dos desafios associados são a privacidade e segurança dos utilizadores, a identificação das intenções do utilizador que interage com o ambiente, a gestão energética e a capacidade de adaptação às alterações na infraestrutura de rede e computação, resultantes do processo evolutivo (Varshney, 2009). A identificação de novos mecanismos de interação entre as diversas tecnologias, a existência de um controlo descentralizado e os tipos de abordagens de desenvolvimento, são também alguns dos mais recentes desafios que se impõem à visão do *pervasive computing* (Conti, 2012).

Contundo, esta visão requer uma alteração de paradigma e a criação de novas abordagens relativamente ao desenvolvimento de *hardware*, redes, *middleware*, serviços e aplicações. O desenvolvimento de tecnologia precisará de ter em conta as diversas dinâmicas entre dispositivos, para ser bem-sucedida no processo de escalar os dispositivos, aplicações e suas utilizações no dia-a-dia (Cook and Das, 2012). O *pervasive computing* pode, assim, ser praticamente aplicado a qualquer aspeto da vida e atividade humana.

Especificamente, a área da saúde tem sido vista como uma das que mais pode ser revolucionada com a implementação de tecnologias *pervasive*. A introdução das tecnologias de informação e comunicação na saúde levou à possibilidade de realizar a atividade de forma mais eficiente e com maior qualidade (Varshney, 2009).

Atualmente, existem muitos desafios, tais como o número significativo de erros médicos (Kohn et al., 2000), o aumento dos custos nos países da Organização para a Cooperação e Desenvolvimento Económico (OCDE) (Zwicker et al., 2011) e o aumento do *stress* dos sistemas de saúde e seus funcionários. Isto tudo, combinado com o aumento da esperança média de vida e custos com a saúde revela os a dificuldade dos desafios existentes (Varshney, 2009).

Nos Estados Unidos da América (EUA) é estimado que 88,5 milhões de pessoas terão mais de 65 anos em 2050 (Vincent and Velkoff, 2010), estando ainda previsto uma queda no número de enfermeiros e médicos para 800 mil e 200 mil, respetivamente (Salzhauer, 2005). Assim,

reconhecem a Fundação Nacional de Ciência e os Institutos Nacionais de Saúde, que as tecnologias de informação e comunicação estão destinadas a transformar a forma como é a informação de saúde acessada e são efetuados os tratamentos clínicos (Cook and Das, 2012). No entanto, os EUA não são um caso isolado e esta situação tem se refletido por todo o mundo, demonstrando assim a necessidade de que os sistemas de saúde terão de ser cada vez mais eficientes, fazendo mais com menos, ou seja, com recursos humanos e financeiros limitados.

As atuais e emergentes tecnologias sem fios poderão aumentar a qualidade geral dos serviços para os doentes, tanto das cidades como de zonas rurais; reduzir o *stress* dos profissionais de saúde, enquanto ao mesmo tempo aumenta a sua produtividade, retenção e qualidade de vida, o que a longo prazo se traduz numa redução dos custos com os serviços de saúde (Varshney, 2009). Muitos dos erros médicos são resultado da falta de informação correta e completa, no local e momento necessário, resultando em diagnósticos errados. O erro médico não será eliminado, mas parte dos erros associados a erros de informação podem sê-lo. Para isso as infraestruturas tem de ser adaptadas às necessidades dos cuidados de saúde (Varshney, 2009). Mais ainda, na área das doenças crónicas acredita-se que através destas tecnologias se pode fornecer ao doente um nível de cuidado próprio com mais qualidade (Wickramasinghe et al., 2013).

Esta visão do *pervasive computing* é, assim, cada vez mais uma realidade. O *pervasive computing* possibilita a criação de novos dados/informações e pode revolucionar quase qualquer aspeto concebível no pensamento, mas para isso é preciso ter em conta as considerações a fazer para poder tornar tal visão realidade.

## **2.8 A Timeline**

A cronologia é a área de estudo que se debruça pela organização de eventos ou datas pela ordem em que aconteceram. É uma atividade tão intrínseca ao nosso ser que, nos dias de hoje, se torna quase impossível imaginar que adquirimos essa habilidade.

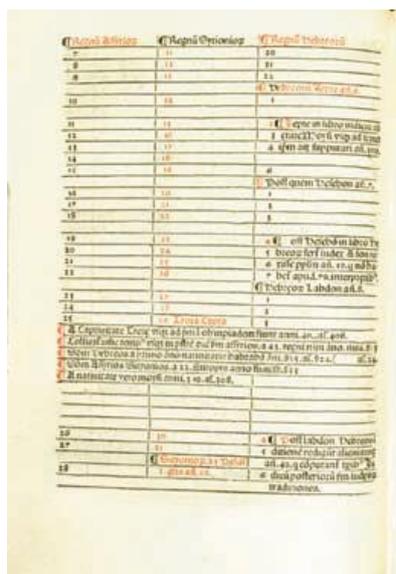
Contudo, o tempo como o representamos hoje nem sempre foi uma realidade. A representação do tempo como é feita hoje em dia foi processo demorado. A nossa capacidade de representar graficamente é das competências mais importantes que desenvolvemos e que nos permite organizar informação, abrindo portas à identificação de padrões, e por conseguinte à aquisição de novo conhecimento.

Para entender o problema da cronologia, é preciso ir até ao conceito da linha e perceber a sua ubiquidade, flexibilidade e força. A linha, seja ela reta ou curva, técnica ou artística, simples ou

embelezada, ramificada ou cruzada é uma figura muito mais complexa e colorida do que parece. Na tentativa de representar o tempo, a linha aparece virtualmente em qualquer sítio, seja em imagens, textos e dispositivos (Rosenberg and Grafton, 2010).

A nossa noção de tempo está de tal forma envolta com a metáfora da linha que nos parece virtualmente impossível separá-los. A maneira como falamos sobre “muito” e “pouco” tempo, “intervalos” de tempo, o “antes” e o “depois”, são tudo metáforas que depende da representação mental do tempo como algo linear. Esta simbiose existente entre a nossa percepção do espaço e noção de tempo é de tal forma forte que não somos capazes de representar o tempo sem a mediação do espaço (Mitchell, 1980).

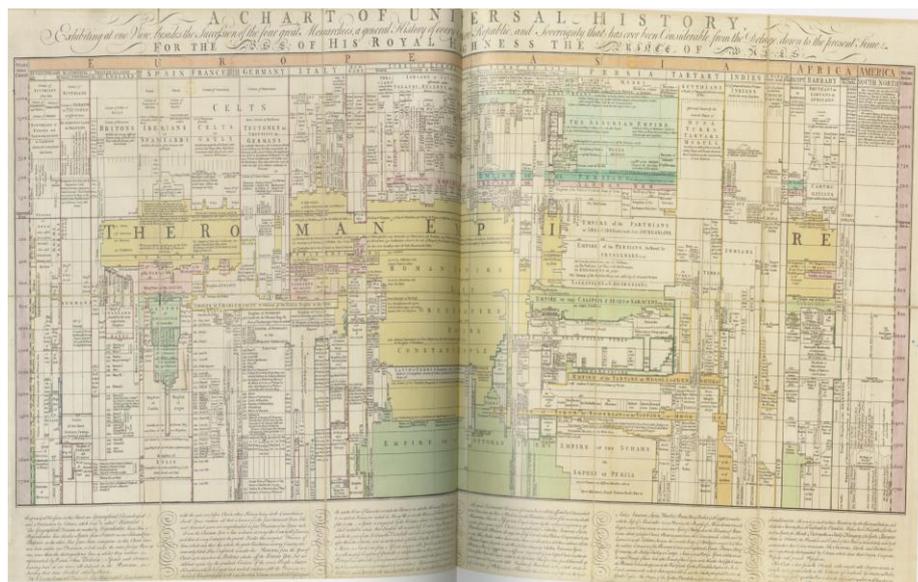
A *Timeline* é uma das várias e mais populares maneiras de representação gráfica de informação. O que a distingue de outras soluções é o facto de apresentar a informação, normalmente eventos ou acontecimentos, ordenada cronologicamente (Rosenberg and Grafton, 2010). Com a *timeline* é possível perceber quando ocorreu um evento, a sua duração e sobreposição com outros eventos. A granularidade associada à linha temporal da *timeline* depende sempre do nível de detalhe que se pretende, seja segundos, dias, anos ou décadas. Esta informação é útil pois permite observar padrões em vários períodos de tempo, graças à forma de apresentar informação da *timeline*. Historicamente, as *timelines* têm início no mundo antigo. No século IV, Eusebius de Caesarea, teólogo cristão, projetou e compôs a *Chronicle* (Figura 4), que se transformou no modelo a seguir para as *timelines* (Rosenberg and Grafton, 2010).



**Figura 4 – A queda de Troia em *Chronicle* de Eusebius (retirado de Rosenberg and Grafton, 2010)**

A apresentação de informação em tabelas revelou ser fácil de reproduzir, permitindo organizar qualquer tipo e quantidade de informação, e facilitar a comunicação entre nações que possuíssem diferentes concepções culturais de tempo.

Nos anos de 1750, Thomas Jefferys desenvolveu o seu gráfico *Chart of Universal History* (Figura 5), que representou o início de uma nova era para a representação do tempo. Isto porque na sua representação não havia informação dividida em células discretas e indexadas. Todo o espaço era uma só representação contínua (Rosenberg and Grafton, 2010).



**Figura 5 – *Chart of Universal History* de Thomas Jefferys (retirado de Rosenberg and Grafton, 2010)**

Mais tarde em 1765, Joseph Priestley através da sua *Chart of Biography*, estabeleceu pela primeira vez o vocabulário para um mapa do tempo, e foi o primeiro a competir com a representação matricial, como uma estrutura normativa para representar a cronologia (Rosenberg and Grafton, 2010).

A partir do século XIX, o uso da *timeline* proliferou e teve diversas aplicações, tornando-se a objetividade cada vez mais uma obsessão trazida pelo aparecer da fotografia (Rosenberg and Grafton, 2010). No entanto, todas estas representações eram estáticas e em papel, tendo limitações ao nível de espaço, alteração do conteúdo e exigindo uma forte perceção prévia de como se iria visualizar a informação (Rosenberg and Grafton, 2010).

Já no século XXI, e com o avanço da tecnologia, é agora possível criar e visualizar *timelines* em computadores, que permitem alterar o conteúdo da *timeline* sem esta ter de ser totalmente refeita, sem limitações de tamanho, criar diferentes categorias, entre outras; tudo isto de uma forma interativa (Richardson et al., 2008). A Figura 6 apresenta um exemplo de uma *timeline* digital, a

Timeline JS desenvolvida pela Northwestern University Knight Lab. Esta *timeline* foi utilizada pela revista TIME como forma de representar a vida de Nelson Mandela, e permite ao utilizador visualizar o conteúdo que esta disponibiliza através de várias funcionalidades, tais como aumentar/diminuir detalhe, avançar/retroceder, entre outras.



**Figura 6 – Exemplo de uma *timeline* digital (retirado de TIME Staff, 2013)**

É mais comum ver a utilização de *timelines* num contexto histórico procurando catalogar acontecimentos, como movimentos sociais, epidemias, guerras e história da vida (Rosenberg and Grafton, 2010), ou ainda o percurso de vida de artistas, o progresso de áreas científicas, entre outros. Mas é também usual em áreas como a Matemática, através da representação gráfica de funções no contexto de um problema, na Sismologia, através dos sismógrafos, na Meteorologia, com representação gráfica da pressão atmosférica ao longo dos dias, na Investigação Criminal, com a criação de cronogramas que retratam a ordem acontecimentos até ao momento do crime, entre outras.

No âmbito organizacional, são também cada vez mais usadas para efetuar planeamento de projetos, onde indicam o início/fim do projeto, tarefas a realizar, quem as realiza, a sua duração, entre outros; sendo que tudo é posteriormente representado num Diagrama de Gantt (Wilson, 2003), o qual ordena as tarefas por ordem temporal e as representam num espaço.

Na área da Medicina, as *timelines* são mais usadas como registos históricos de surtos de doenças, estudos sobre doenças, ou na prática diária de medicina servem a representação de dados clínicos. Em Medicina Intensiva, são geralmente usadas, e com monitorização contínua, em eletrocardiogramas, monitorização de CO<sub>2</sub> expirado, monitorização de pressão sanguínea por

cateter central, entre outros (McDonald and Gardner, 1987). No entanto, o pretendido com este projeto é implementar as *timelines* de uma forma inovadora nas Unidades de Cuidados Intensivos (UCI). Através do desenvolvimento de uma *Pervasive Patient Timeline* interativa, pretende-se que os profissionais de saúde tenham acesso ao historial clínico dos doentes, desde que estes foram internados até ao momento atual, sendo possível adicionar também previsões futuras. O acesso ao ambiente será em tempo real, apresentando os dados recolhidos no momento, sendo também possível consultar eventos passados, como medicação, análises, sinais vitais, entre outros. A apresentação de registos históricos surge com o objetivo de prever o estado futuro dos doentes através da combinação com técnicas de DM, por forma a ajudar os profissionais de saúde a criar diagnósticos mais precisos, que ajudem na escolha do tratamento mais adequado para os doentes.

Uma das características que distingue esta *timeline* é o facto de ser *pervasive*, preconizando a automatização de processos, o acesso em tempo real à informação, a qualquer hora e em qualquer lugar, bem como o envio de alertas no caso de alterações significativas no estado de saúde dos doentes. Além disso, é ainda possível apresentar os resultados dos modelos de *Data Mining* (DM) do sistema INTCare na *timeline*, através da apresentação da probabilidade de acontecer um determinado evento clínico nas próximas horas (ex: toma de vasopressores).

Esta aplicação das *timelines* em Medicina Intensiva é uma abordagem inovadora que facilita a perceção da informação por parte dos profissionais de saúde, ajuda a desenvolver relação entre eventos e acima de tudo promove a prevenção de estados clínicos.

## Capítulo 3 – Abordagem Metodológica

### 3.1 Metodologia de Investigação

A abordagem metodológica seguida durante a realização deste projeto de dissertação foi o *Design Science Research* (DSR). O DSR tem sido uma abordagem que se encontra em rápido desenvolvimento, tendo inclusive acrescentado a palavra '*Science*' ao que era até então denominado *Design Research* (DR). Apesar do DR ser uma abordagem que podia ser aplicada a diversas áreas, não possuía a característica que o DSR disponibilizava: a aprendizagem através da construção. É também usado o argumento de que o DR é a pesquisa 'para' ou 'sobre' projetar, enquanto o DSR é pesquisa usando a capacidade de projetar como método de pesquisa. Assim, quando definido como aprendizagem através da construção, o DSR não é único aos sistemas de informação. As áreas da educação, saúde, ciências da computação e engenharia também são exemplos da sua utilização (Vaishnavi and Kuechler, 2004).

Esta metodologia consiste num conjunto de técnicas de análise e perspetivas para realizar investigação em Sistemas de Informação. É constituída por seis atividades, que constituem o processo de iteração (Figura 7). O momento de perceção do problema é dividido em duas atividades: Identificar Problema e Motivar e Definir Objetivos da Solução; o desenvolvimento da solução diz respeito à atividade, Projetar e Desenvolver; o momento de teste e avaliação é dividido, respetivamente, nas atividades Demonstração e Avaliação; e por fim à atividade final é atribuído o nome de Comunicação (Peffer et al., 2007).

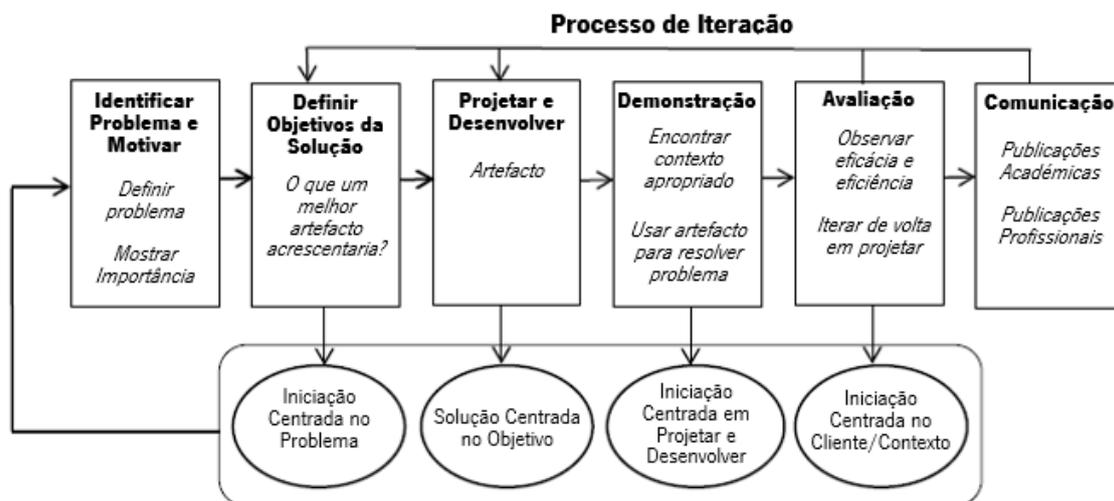


Figura 7 – Modelo de Processo da Metodologia *Design Science Research* (adaptado de Peffer et al., 2007)

Na atividade Identificar Problema e Motivar, deve ser definido o problema específico em análise e apresentada uma justificação referente ao valor da solução. O problema deve ser conceitualmente atomizado, por forma a permitir que a sua complexidade seja capturada pela solução. A justificação do valor da solução dá motivação ao investigador para procurar a solução e aceitar os resultados, e ajuda-o a ter uma melhor perceção sobre o problema que procura resolver. A atividade Definir Objetivos da Solução procura inferir os objetivos da solução, através da definição do problema e do conhecimento do que é possível fazer. Para a atividade Projetar e Desenvolver está planeada a criação do artefacto. Na atividade Demonstração, deverá ser demonstrado o uso do artefacto produzido para resolver uma ou mais instâncias do problema. Já a atividade de Avaliação conclui que se deve observar e medir a capacidade de o artefacto satisfazer a solução definida para o problema. Por último, a atividade de Comunicação admite que se deve comunicar o problema e a sua importância, o artefacto, a sua utilidade, o grau de inovação, rigor da sua projeção e eficácia em resolver o problema (Peppers et al., 2007).

O modelo permite que este seja iniciado em qualquer uma das atividades, exceto a final; dependendo apenas do tipo de orientação que se pretende dar ao processo (Peppers et al., 2007). Esta metodologia procura criar conhecimento através do desenvolvimento de novos ou inovadores artefactos, bem como analisar e avaliar a performance e eficácia desses mesmos artefactos, sempre com o foco em melhorar e compreender os diferentes aspetos dos Sistemas de Informação (Vaishnavi and Kuechler, 2004).

No contexto deste trabalho, esta metodologia foi aplicada tanto no desenvolvimento da *Pervasive Patient Timeline*, como na criação de modelos de *Data Mining* (DM) para prever a administração de vasopressores.

Em ambos os momentos, o processo de iteração teve sempre um início centrado no problema, tendo-se iniciado no momento em que foram identificados o problema e a importância da sua resolução. Posto isto foi iniciada a primeira atividade Definir Problema e Motivar, e desta forma no caso da *Pervasive Patient Timeline* o problema esteve em perceber de que forma o seu desenvolvimento poderia contribuir para a melhoria do processo de decisão. Já no caso da criação de modelos de DM, o problema observado diz respeito à possibilidade de usar técnicas de DM para prever a probabilidade de um doente vir a precisar de tomar vasopressores.

No que diz respeito à relevância em ver estes problemas respondidos, esta prende-se com a importância que esta previsão pode ter em diminuir os erros médicos, ajudando os profissionais de saúde a tomar decisões mais acertadas, para o bem dos doentes.

A segunda atividade da metodologia, Definir Objetivos da Solução, focou-se na definição dos objetivos. No capítulo da *Pervasive Patient Timeline* foi definido como objetivo o desenvolvimento de uma *pervasive timeline* interativa capaz de providenciar novo conhecimento de forma a facilitar os processos de tomada de decisão em ambientes clínicos, sem restrições de local e tempo. Por sua vez, com a criação de modelos DM, o objetivo definido foi perceber se seria possível através do uso de técnicas de DM prever a probabilidade de um doente vir a necessitar de tomar vasopressores.

Na atividade seguinte, Projetar e Desenvolver, para a *Pervasive Patient Timeline* foram executadas as tarefas de projeção das características do artefacto, culminando com o desenvolvimento do artefacto. Relativamente aos modelos de DM, as tarefas executadas estão associadas com a compreensão, preparação dos dados e criação dos modelos de DM.

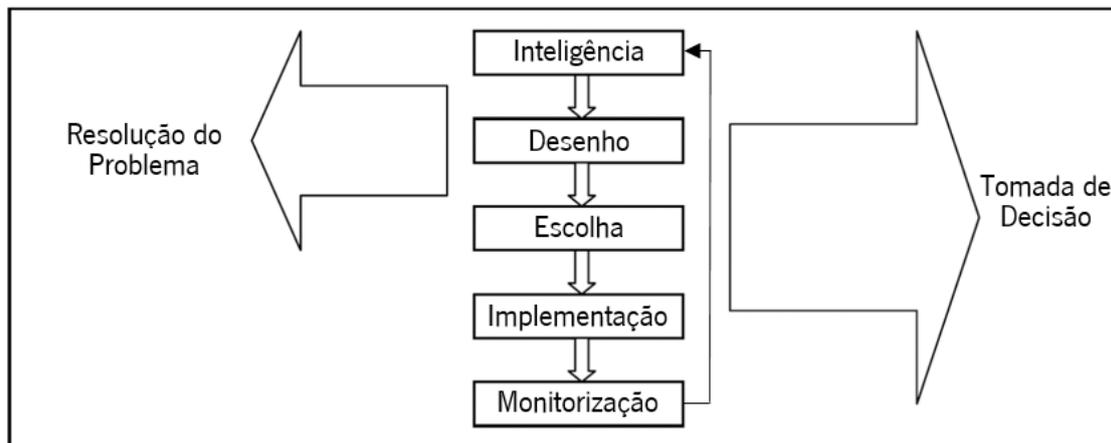
Já na fase de Demonstração, para a *Pervasive Patient Timeline* foram incorporados dados clínicos de forma a simular a disposição e apresentação da informação na timeline. Para os modelos de DM, a demonstração concretiza-se na aplicação dos modelos desenvolvidos a outros doentes em condições semelhantes.

No desenvolvimento da *Pervasive Patient Timeline*, a atividade Avaliação consistiu na criação de um conjunto de métricas que poderão ser usadas para avaliar a performance da *timeline*. No que toca aos modelos de DM, esta atividade focou-se em avaliar os resultados obtidos pelos modelos e se estes satisfazem os objetivos propostos.

Por último, a comunicação do trabalho desenvolvido foi efetuada, em ambos os casos, através da escrita de artigos científicos publicados em conferências e revistas.

### **3.2 Modelo do Processo de Tomada de Decisão**

Para o desenvolvimento da *Pervasive Patient Timeline* foi usado o Modelo do Processo de Tomada de Decisão (MPTD) proposto por Herbert Simon (1960) para o desenvolvimento de sistemas de apoio à decisão (Figura 8).



**Figura 8 – Modelo do Processo de Tomada de Decisão (adaptado de Tripathi, 2011)**

Simon caracterizou três fases no seu modelo: Inteligência, Desenho e Escolha. Para Simon a fase Implementação estaria incluída na fase Escolha, no entanto vários autores acham que a fase de implementação tem relevância e importância suficiente para ser representada separadamente (Sprague and Ralph, 1980). É transversal a qualquer processo de tomada de decisão, que se deve começar por reconhecer o problema a resolver. A fase Inteligência caracteriza-se pela análise, exploração e descrição do problema. Procura-se no ambiente por condições que impliquem decisões. Dados são recolhidos, processados e examinados, na senda de encontrar pistas que possam identificar problemas. Em seguida, vem a fase Desenho, onde se inventam, desenvolvem e analisam possíveis cursos de ação. Isto envolve, a execução de processos para perceber o problema, desenvolvimento de soluções e testes, a fim de avaliar a viabilidade das mesmas. Na fase Escolha, procura-se escolher o melhor caminho, de entre os propostos na fase anterior. A fase Implementação é onde ocorre a implementação do artefacto desenvolvido, depois de esta cumprir os requisitos e respetivos testes de avaliação (Sprague and Ralph, 1980). Por último, vem a fase de Monitorização, que visa avaliar, monitorizar e validar o desempenho do artefacto desenvolvido, já depois de este estar implementado (Turban et al., 2014). Este modelo ainda que comumente aceite, é apenas mais um modelo entre vários que pretendem dar suporte ao processo de tomada de decisão, dado a inexistência de um modelo aceite como *standard* (Sprague and Ralph, 1980).

No âmbito deste trabalho foram desempenhadas o conjunto de fases que o modelo aborda, tendo estas se adaptado ao contexto da solução em desenvolvimento.

Na fase de Inteligência foi identificado o problema a resolver e o contexto em que este está inserido. A partir daqui foram também definidos os objetivos a atingir com o desenvolvimento de uma

solução para o problema e o processo de pesquisa usado para procurar soluções. De seguida, foi iniciada a fase de Desenho, onde foram definidos requisitos mínimos que as soluções a procurar deveriam ter. Foram de seguida procuradas soluções que satisfizessem os requisitos propostos, tendo sido explicado o contexto de cada uma. Posteriormente, foi executada a fase Escolha na qual são apresentadas as razões da escolha de uma das soluções em detrimento das outras. Na fase de Implementação foi descrito o processo de desenvolvimento da solução, as tecnologias e ferramentas usadas, as linguagens de programação, bem como o conjunto de funcionalidades presentes na solução.

Por último, na fase de Monitorização foram definidas métricas que permitam analisar a performance da solução quando esta estiver completamente implementada no sistema INTCare.

### 3.3 Metodologia de *Data Mining*

O *Cross Industry Standard Process for Data Mining* (CRISP-DM) surge como uma metodologia capaz de ajudar a conduzir o processo de DCBD, na medida em que define um projeto como um processo cíclico, em que através da execução de várias iterações se chega a um resultado final que vai mais ao encontro dos objetivos definidos para o projeto (Moro et al., 2011). O CRISP-DM é assim constituído por seis fases: Compreensão do Negócio, Compreensão dos Dados, Preparação dos Dados, Modelação, Avaliação e Implementação. A Figura 9 ilustra o processo da metodologia:

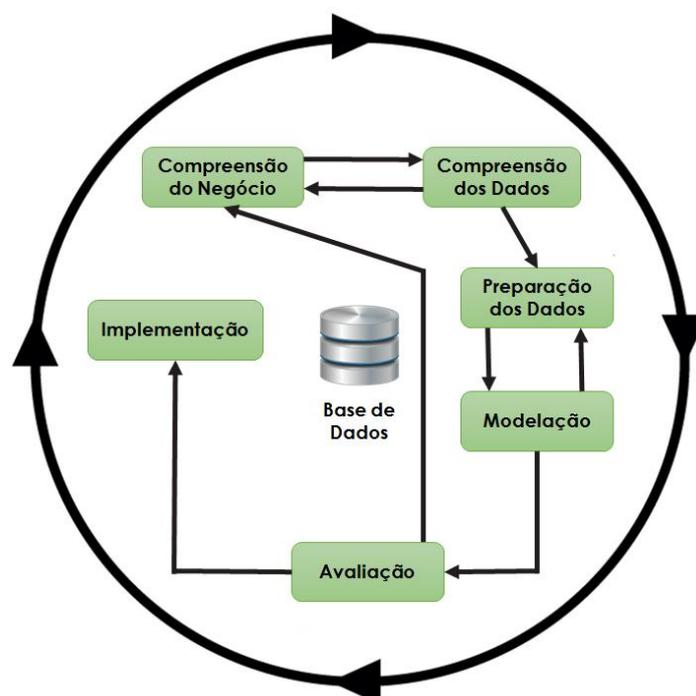


Figura 9 – Ciclo de vida do CRISP-DM (adaptado de Marcelino et al., 2015)

A primeira fase (Compreensão do Negócio) tem como objetivo entender o negócio, através da clarificação dos conceitos a ele associados e a definição dos objetivos. Na segunda fase (Compreensão dos Dados) é feita uma exploração e análise dos dados, de forma a perceber a sua qualidade, bem como significado. A terceira fase (Preparação dos Dados) foca-se em tratar os dados, corrigindo ou eliminando erros existentes, criando novos dados através de agregação, entre outros. Estas operações têm o intuito de criar um conjunto final de dados mais simples e de maior qualidade para mais tarde ser processado por técnicas de DM. Na quarta fase (Modelação) são selecionadas e posteriormente aplicadas as técnicas de DM ao conjunto de dados em análise, sendo assim construídos modelos de DM. A quinta fase (Avaliação) é o momento em que se avalia a qualidade de previsão dos modelos construídos. Caso os resultados dos modelos vão ao encontro do desejado passa-se então à sexta, e última, fase (Implementação), onde os modelos são integrados no ambiente de trabalho, caso contrário deve-se reiterar todo o processo até se estar satisfeito. O CRISP-DM não é uma metodologia rígida, pelo que permite o retrocesso nas fases que compõem o seu ciclo de vida, de forma a melhorar a qualidade do trabalho desenvolvido (Marcelino et al., 2015).

No domínio do trabalho desenvolvido, o CRISP-DM foi aplicado com o objetivo de orientar o desenvolvimento do trabalho. Na fase de Compreensão do Negócio foi descrito o problema encontrado, o contexto do problema a resolver, os objetivos a atingir e tecnologias utilizadas. A seguinte fase é a Compreensão dos Dados, onde foi efetuada uma análise ao conjunto de dados usado para os modelos de DM. Na fase de Preparação dos Dados foram executadas diversas tarefas de correção do estado dos dados, preparando-os para um estado de qualidade desejado. Para a fase de Modelação foram desenvolvidos modelos de DM que atuam sobre o conjunto de dados, e produzem resultados. Na fase de Avaliação foram avaliados os resultados adquiridos, através das curvas ROC e das métricas obtidas a partir da matriz de confusão. Por fim na fase de Implementação foram integrados os modelos de DM desenvolvidos na *Pervasive Patient Timeline*, podendo assim ser possível observar futuras previsões para os doentes (ex: toma de vasopressores).

### **3.4 Cruzamento entre DSR e MPTD**

Como forma de assegurar o bom desenvolvimento tanto do trabalho de investigação como o trabalho de desenvolvimento do artefacto *timeline*, tanto o DSR como o MPTD foram usados em simultâneo, retirando-se assim vantagens da sinergia resultante. Na Tabela 3, são apresentadas

as atividades e fases do DSR e MPTD, respetivamente. Mais se apresenta o cruzamento das atividades e fases que se identificam umas com as outras.

**Tabela 3 – Cruzamento DSR vs MPTD**

	<b>DSR</b>	<b>MPTD</b>
<b>Fases</b>	Identificar Problema e Motivar	Inteligência
	Definir Objetivos da Solução	
	Projetar e Desenvolver	Desenho Escolha
		Implementação
	Demonstração	
	Avaliação	Monitorização
	Comunicação	

As atividades Identificar Problema e Motivar e Definir Objetivos da Solução da metodologia DSR coincidem com a fase Inteligência do modelo MPTD, isto porque representam o momento inicial de identificação e perceção do problema, bem como a definição de objetivos.

A atividade Projetar e Desenvolver corresponde à fase Desenho, Escolha e Implementação, pois trata-se de criar uma ideia e modelo das funcionalidades esperadas para o artefacto, encontrar soluções adequadas, escolher a melhor entre as possíveis e por fim implementá-la.

A fase de Implementação do MPTD identifica-se também com a atividade Demonstração do DSR, visto ser o momento em que se procura utilizar o artefacto e aplicá-lo a um contexto real, tendo neste caso sido feito através da incorporação de vários dados clínicos dos doentes no artefacto. A última relação existente é entre a atividade Avaliação e a fase Monitorização, onde ambas pretendem avaliar a qualidade e desempenho do artefacto já implementado. A atividade Comunicação não revela ter relação com nenhuma atividade do MPTD, isto porque o resultado da sua execução é único e apropriado ao DSR. A atividade Comunicação, tendo um carácter mais científico e de investigação, pretende comunicar o artefacto, as suas vantagens, que problema é que vem resolver e qual a importância que existe em que esse problema seja resolvido.

### 3.5 Cruzamento entre DSR e CRISP-DM

Para o desenvolvimento dos modelos de DM foi usada a metodologia CRISP-DM em simultâneo com o DSR. Na Tabela 4 é possível observar que atividades e fases de ambas as metodologias se cruzam no desenrolar do ciclo de vida de cada uma.

**Tabela 4 – Cruzamento DSR vs CRISP-DM**

	<b>DSR</b>	<b>CRISP-DM</b>
<b>Fases</b>	Identificar Problema e Motivar	Compreensão do Negócio
	Definir Objetivos da Solução	
	Projetar e Desenvolver	Compreensão dos Dados
		Preparação dos Dados
	Demonstração	Modelação
	Avaliação	Avaliação
		Implementação
	Comunicação	

O primeiro cruzamento existente entre as duas metodologias ocorre com as atividades Identificar Problema e Motivar e Definir Objetivos da Solução do DSR e a fase Compreensão do Negócio do CRISP-DM. Em ambas o foco está em perceber a situação em análise, através da identificação do problema e dos objetivos propostos para o resolver. A segunda situação de correspondência entre o DSR e CRISP-DM verifica-se entre a atividade Projetar e Desenvolver e as fases Compreensão dos Dados, Preparação dos Dados e Modelação. Aqui para o DSR o foco está na realização de passos que levem à construção do artefacto. Isto coincide com as referidas fases do CRISP-DM pois é nessas fases que são realizadas tarefas mais técnicas que irão culminar no artefacto, neste caso a modelação de modelos de DM.

De seguida, há também cruzamento entre a atividade Demonstração do DSR e a fase Modelação do CRISP-DM. Tanto numa como outra, o objetivo passa por introduzir o artefacto num ambiente com problemas semelhantes aos que se espera que este resolva. Esta aplicação ocorre quando os modelos de DM desenvolvidos são aplicados a novos doentes em situações semelhantes.

Por fim, existe simultaneidade entre a atividade de Avaliação do DSR e a fase Avaliação do CRISP-DM. O objetivo de ambas é avaliar se o artefacto desenvolvido, neste caso modelos de DM, satisfaz os objetivos propostos, ou seja, se tem a qualidade esperada.

A atividade Comunicação não tem qualquer correspondente na metodologia oposta, isto porque não se identifica adequadamente com os propósitos das fases existente



## **Capítulo 4 – Desenvolvimento da Solução**

### **4.1 Inteligência**

A inteligência é a primeira fase do Modelo do Processo de Tomada de Decisão (MPTD) sendo constituída por um conjunto de atividades que pretendem preparar o desenvolvimento do artefacto final.

A necessidade do desenvolvimento desta solução foi detetada por intensivistas da Unidade de Cuidados Intensivos (UCI) do Centro Hospitalar do Porto (CHP). No âmbito do projeto INTCare, perspectiva-se que a solução possa integrar alguns dos resultados atingidos pelo projeto principal. O problema identificado diz respeito à dificuldade sentida pelos intensivistas em analisar informação proveniente de diversas e variadas fontes de informação, que podem possuir diferentes formas de apresentação do conteúdo. Desta forma o objetivo da solução é suprimir esta dificuldade existente através da implementação de um modelo de *timeline* capaz de dar apoio às decisões dos intensivistas.

Para o desenvolvimento deste artefacto foram pesquisadas várias informações associadas a palavras como “*timeline*”, “*software*”, “*dynamic*”, “*visualization*”, “*plugins*”, “*widget*”, que através de combinações geraram resultados de *timelines*.

### **4.2 Desenho**

Esta fase foi iniciada com a procura de *timelines* que pudessem ser opção para o desenvolvimento da *Pervasive Patient Timeline*. Desta forma, foram definidos alguns requisitos mínimos que se esperava que as *timelines* tivessem com o intuito de identificar algo que servisse como base de implementação e que se enquadrasse no tempo de desenvolvimento do projeto. Os requisitos inicialmente definidos foram:

#### **Dados**

- Possibilidade de carregar na *timeline* dados provenientes de bases de dados com diversas fontes de informação.

#### **Granularidade**

- Máximo: 15 minutos;
- Mínimo: Década.

## Usabilidade

- Arrastar a *timeline* na horizontal, por forma a visualizar eventos passados e futuros;
- Aumentar e diminuir a granularidade enquanto se visualiza a *timeline*;
- Filtro para restringir categorias visíveis na *timeline*.

## Características dos Eventos

- Título;
- Descrição do evento;
- Eventos com data de início e data de fim;
- Capacidade de adicionar eventos a diferentes categorias.

Assim com a definição destes requisitos foi iniciada a procura por *timelines*, tendo sido encontradas três *timelines* que apresentaram capacidades de se adaptar e poder incorporar os requisitos necessários: a SIMILE Timeline, a Vis.js e a Timeglider JS.

## SIMILE Timeline

A SIMILE Timeline (Figura 10) é uma *timeline* resultante do projeto SIMILE desenvolvido no Massachusetts Institute of Technology (MIT), que é agora mantida por uma comunidade de desenvolvedores *open-source*.

Esta *timeline* revelou ser capaz de cumprir todos os requisitos necessários, apresentando diversos exemplos em que a *timeline* foi usada de forma a corroborar essas capacidades. Esta solução apresentou extensa documentação sobre a *timeline* com tutoriais de implementação

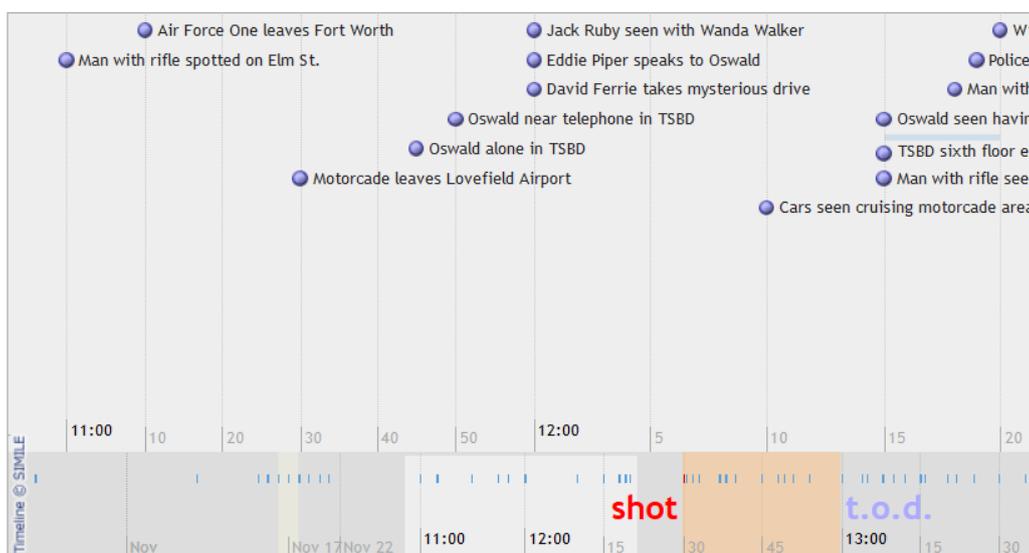


Figura 10 – Exemplo da SIMILE Timeline (retirado de SIMILE Project)

## Vis.js

A Vis.js (Figura 11) é uma *timeline* desenvolvida em *Javascript*. Esta *timeline* apresentou ser capaz de cumprir com a maioria dos requisitos, tendo uma boa apresentação ao nível das funcionalidades com exemplos para as testar, apresentação da documentação com a descrição de métodos, eventos, entre outros. Contudo não continha informação relativamente à capacidade de incorporar dados provenientes de bases de dados.



Figura 11 – Exemplo da Vis.js (retirado de Vis.js)

## Timeglider JS

A Timeglider JS (Figura 12) é uma *timeline* desenvolvida em *Javascript* e que apresentou ser capaz de cumprir com todos os requisitos pressupostos. Apresentou ter os requisitos mínimos exigidos com uma agradável implementação e apresentação no interface. Esta *timeline* apresentou vários exemplos das suas funcionalidades, continha boa documentação sobre a *timeline*, inclusive com tutoriais de implementação.

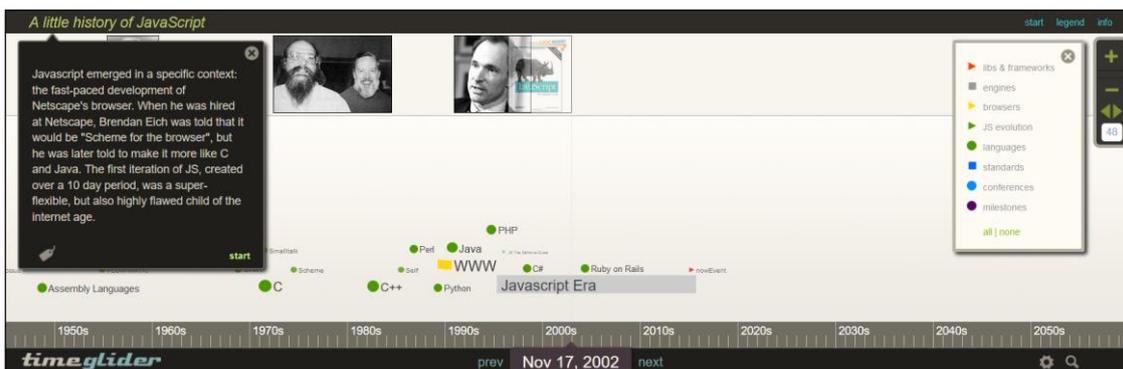


Figura 12 – Exemplo da Timeglider JS (retirado de Timeglider JS)

Com base nas *timelines* identificadas é de esperar conseguir tirar partido das bases de implementação por elas fornecidas, por forma a desenvolver uma *timeline* que quando inquirida seja capaz de apresentar os dados do doente requisitado com as funcionalidades definidas como requisitos. A *timeline* deverá também ser capaz de visualizar dados passados e futuros do doente,

através da incorporação de modelos de *Data Mining* (DM), divisão dos dados por categorias tais como sinais vitais, análises clínicas, entre outras.

### 4.3 Escolha

De entre as três *timelines* a que mais se mostrou promissora foi a Timeglider JS. As razões principais que levaram à escolha desta *timeline* face às outras duas estão relacionadas com a facilidade do desenvolvimento da *timeline*, funcionalidades extras e apresentação do conteúdo na *timeline*. A Timeglider JS foi a que apresentou ser a mais compacta das três visto já ter parte dos requisitos incorporados, tornando desta forma mais reduzido o tempo de desenvolvimento para adaptar o sistema aos requisitos definidos. Isto é visível ao nível da divisão por categorias, as quais já fazem parte do sistema. O carregamento de conteúdo vindo de bases de dados para a *timeline* foi o mais fácil de entender face à existência de exemplos bem definidos. Na SIMILE a forma de carregar os dados pareceu um pouco mais confusa, e na Vis.js não foi encontrada documentação relativa a carregamento via bases de dados.

Ao nível de funcionalidades extra revelou ser possível adicionar informações relativamente à *timeline*, navegar de evento para evento sem necessidade de arrastar quando os eventos estão mais afastados uns dos outros, usar um filtro de procura de palavras na *timeline*, definir que eventos aparecem na *timeline* tendo em conta o nível da granularidade e também a possibilidade de adaptar as datas dos eventos conforme o fuso-horário. Já nas restantes *timelines* não foram detetadas funcionalidades extra já integradas no sistema.

No que toca ao interface, este revelou ser mais funcional, mais desenvolvido, com melhor aspeto ao nível da apresentação dos eventos na *timeline*, boa facilidade de leitura, capacidade de apresentar imagens relativas aos eventos, endereços eletrónicos e boa capacidade de resposta à existência de vários eventos presentes ao mesmo tempo na *timeline*. A Vis.js revelou a necessidade de integrar várias funcionalidades dispersas, para se construir um interface com todos os requisitos e a SIMILE apesar de já ter um interface mais composto, ainda havia necessidade de integrar algumas funcionalidades (ex: categorias).

Como forma de ajudar na classificação das *timelines*, foi feito um cruzamento entre determinadas características e as *timelines* pesquisadas. O cruzamento procura clarificar a perceção adquirida sobre a capacidade de cada uma das *timelines*, justificando assim o que levou à escolha da Timeglider JS como base de desenvolvimento. Na Tabela 5 é possível verificar o cruzamento das *timelines* com os requisitos mínimos, funcionalidades extras existentes, qualidade do interface,

tempo de desenvolvimento, qualidade da documentação e a classificação final. Para o interface, tempo e documentação foi atribuído um nível de 1 a 5, sendo 1 o pior e 5 o melhor. A análise constitui o parecer de três pessoas que observaram a timeline.

**Tabela 5 – Cruzamento das *timelines* e suas características**

	<b>Requisitos Mínimos</b>	<b>Extras</b>	<b>Interface (1-5)</b>	<b>Tempo (1-5)</b>	<b>Documentação (1-5)</b>	<b>Final</b>
<b>SIMILE</b>	Sim	Não	3	3	4	2°
<b>Vis.js</b>	Não	Não	1	2	2	3°
<b>Timeglider JS</b>	Sim	Sim	5	4	4	1°

No geral a Timeglider JS apresentou os requisitos pedidos já integrados, com melhor apresentação, boa documentação e ainda com funcionalidades extras. Tudo isto acabou por contribuir para um tempo de desenvolvimento mais reduzido.

Definida a tecnologia para base do desenvolvimento da *Pervasive Patient Timeline*, foram delineadas orientações a seguir, por ordem crescente, durante o processo de implementação.

- 1º** - Definir tecnologias e ferramentas a utilizar para o desenvolvimento;
- 2º** - Identificar como funciona o carregamento de dados vindos de bases dados;
- 3º** - Definir dados para o carregamento da *timeline*;
- 4º** - Definir base de dados de doentes para carregar a *timeline*;
- 5º** - Efetuar transformações dos dados da base dados para os incorporar na *timeline*;
- 6º** - Definir categorias dos dados na *timeline*;
- 7º** - Incorporar mais dados (ex: previsões) para finalizar o protótipo.

Este conjunto de orientações foi definido como um caminho a seguir capaz de facilitar o desenvolvimento na fase de implementação. Para além das orientações definidas, não foi descartada a possibilidade de execução de novas orientações e a existência de imprevistos que pudessem impedir a execução das orientações na ordem proposta ou até a execução das mesmas.

#### **4.4 Implementação**

Para a fase de implementação foi desenvolvido um protótipo da *Pervasive Patient Timeline*. O processo de desenvolvimento iniciou-se com a realização de tarefas com a finalidade de definir as

tecnologias e ferramentas a utilizar, perceber o funcionamento geral da *timeline*, as interações entre as diferentes funcionalidades e de que forma poderiam ser acrescentadas novas funcionalidades.

Dada a necessidade de utilizar bases de dados para carregar os dados e de programação web para o desenvolvimento, o protótipo foi construído através da plataforma *XAMPP*.

O *XAMPP* é uma plataforma desenvolvida pela *Apache Friends*, e consiste na agregação de um servidor web *Apache*, uma base de dados *MySQL*, e interpretadores das linguagens de programação *PHP* e *Perl*. O uso desta plataforma permitiu desenvolver e testar o protótipo num só computador, com ou sem internet, isto porque a plataforma agrega num só pacote várias tecnologias, que por norma teriam de ser interligadas separadamente (XAMPP, 2015). Para o tratamento dos dados a carregar foi utilizada a ferramenta *Oracle SQL Developer* que é um ambiente de desenvolvimento integrado para gerir bases de dados. A razão deveu-se ao facto de a base de dados do Centro Hospitalar do Porto estar desenvolvida em *Oracle*.

Ao nível dos dados foi definido que estes deveriam pertencer a três categorias: sinais vitais, análises clínicas e ventilação, sendo apresentados todos os dados ou apenas os críticos. Os dados deveriam ter várias formas de representação: por hora, o primeiro, médio, máximo e mínimo valores atingidos. Também foi definido que deveria ser possível visualizar informação relacionada com a admissão do doente, bem como previsões resultantes de modelos *Data Mining* (DM) em ação no sistema INTCare.

Assim foi preparado um conjunto de dados de doentes, com dados relativos aos sinais vitais, análises clínicas e ventilação, com o intuito de testar a sua introdução na *timeline*.

De seguida foi feita uma análise à *timeline*, onde foram definidas como linguagens de programação web: o *PHP* e *Javascript* para a construção do interface e funcionalidades da *timeline*, e a linguagem *SQL* para a manipulação de dados contidos em bases de dados.

Terminada a escolha de ferramentas e linguagens foi iniciado o desenvolvimento.

## **Desenvolvimento**

O processo inicial passou por experimentar e adaptar exemplos da *timeline* já existentes de forma a perceber o objetivo do código de programação executado e que efeitos tinha na *timeline* a sua modificação. Neste momento foi identificada a forma de carregar diferentes conjuntos de dados na *timeline*, a qual ocorre através do carregamento de um ficheiro que interpreta código *JSON* e o carrega assim que a *timeline* é iniciada, apresentando os eventos.

Posto isto, foi criada uma conexão à base de dados do CHP para poder ser feito o carregamento de dados vindo de base de dados, tendo sido usado um conjunto de dados preparado para ser testado na *timeline*.

O carregamento é feito através de um ficheiro *PHP* que interpreta o conteúdo como sendo *JSON*. O ficheiro incorpora código *PHP* que define as condições de execução do código, efetua a conexão à base de dados, executa *queries SQL* e imprime os dados, tudo isto estruturado de acordo com a sintaxe *JSON* para que depois possa ser interpretado. O processo foi mais tarde refinado de forma a apresentar por hora o primeiro valor, o máximo, o médio e o mínimo.

Após a identificação e sucesso no carregamento de dados, o foco foi orientado para a possibilidade de escolher um doente de uma lista e carregar todos os seus dados ou apenas os críticos. Estas duas funcionalidades foram construídas na página principal onde também está presente o interface da *timeline*. O processo passou pela criação de duas *queries SQL*, uma que retorna o número de processo de doentes, formando a lista de doentes, e outra *querie* cujo objetivo é procurar nos registos de sinais vitais do doente escolhido, aqueles com valor crítico positivo e negativo, no caso de se pretender listar todos os dados, ou apenas valor crítico positivo, no caso de o objetivo ser ver apenas os dados críticos do doente. Este processo utiliza um conjunto de outros procedimentos já definidos, os quais tem como objetivo detetar se um dado é ou não crítico. Em seguida foi tirado partido da capacidade da *timeline* associar os eventos a categorias, pelo que foram feitas as alterações necessárias (processo ETL) para ser possível a introdução das categorias desejadas.

A apresentação dos dados na *timeline* ocorre sempre que haja pelo menos uma das categorias com dados. Isto significa que os intensivistas não precisam de esperar que hajam todos os dados disponíveis, podendo assim visualizar os dados que existem até ao momento, e quando os restantes estiverem prontos passam a aparecer na *timeline*.

De resto foram ainda feitas várias correções ao nível do funcionamento das funcionalidades existentes pois continham erros que resultavam em ações e interações inesperadas. Como exemplo disso é o facto de ter sido verificado que o interface da *timeline* era carregado de cada vez que era feita a atualização dos dados para um novo doente. Há medida que isto acontecia a *timeline* ia como que se repetindo por cima dela própria, sem apagar a instância anterior, ficando cada vez mais lenta a sua utilização e inteligível o conteúdo. Isto foi corrigido de forma que a *timeline* fosse carregada apenas uma vez, sendo apenas substituídos os dados apresentados. A

correção desta situação criou erros noutras funcionalidades (ex: categorias, procura de eventos, fuso-horário,...) que tiveram que ser também corrigidos.

De uma forma geral, para além dos ficheiros que têm como propósito construir a *timeline*, há um ficheiro que é onde o interface da *timeline* aparece e outro que se encarrega de processar os dados e apresentá-los na *timeline*.

Sendo assim após o estudo, exploração e desenvolvimento foram incorporadas as seguintes funcionalidades na *Pervasive Patient Timeline*:

### **Dados**

- Possibilidade de carregar na *timeline* dados provenientes de bases de dados de diversas fontes de dados.

### **Granularidade**

- Máximo: 15 minutos;
- Mínimo: Década.

### **Usabilidade**

- Escolher de uma lista de doentes, o doente que se quer visualizar os dados;
- Escolher o tipo de dados a visualizar: todos os dados ou apenas dados críticos;
- Arrastar a *timeline* na horizontal, por forma a visualizar eventos passados e futuros;
- Aumentar e diminuir a granularidade enquanto se visualiza a *timeline*;
- Filtrar as categorias visíveis na *timeline*;
- Ajustar o tempo de acordo com o fuso-horário;
- Procurar eventos por palavras no título ou descrição;
- Clicar para avançar/retroceder para o evento mais perto;
- Definir uma data e avançar/retroceder para o momento definido.

### **Características dos Eventos**

- Título, valores absolutos da primeira hora e os valores máximo, médio e mínimo da hora;
- Descrição do evento, que pode conter texto, imagens, endereços *web*;
- Eventos com data de início e data de fim;
- Capacidade de adicionar eventos a diferentes categorias;
- Um indicador chamado nível de importância que define a granularidade a que os eventos ficam visíveis (importante se houver vários eventos ao mesmo tempo na *timeline*).

Como resultado do trabalho desenvolvido, a Figura 13 apresenta um exemplo da visualização dos dados de um doente na *timeline*, com o propósito de entender de que forma estão distribuídas as

funcionalidades pelo interface da *timeline*. A roxo e em maioria pode-se observar os dados de monitores de ventilação, a azul os dados de sinais vitais e a verde os dados de análises clínicas. De salientar que para os sinais vitais estão presentes dois valores críticos representados a vermelho. Estão ainda ilustrados e numerados os locais onde se pode aceder a algumas das funcionalidades implementadas:

- 1 – Lista de doentes, dos quais se seleciona um;
- 2 – Escolha entre a visualização de todos os dados ou apenas dados críticos;
- 3 – Descrição de informação sobre a admissão do doente;
- 4 – Lista das categorias existentes e onde se restringe o que aparece na *timeline*;
- 5 – *Slider* de aumento ou diminuição da granularidade da *timeline*;
- 6 – Clicar para definir uma data e avançar para esse momento;
- 7 – Avanço/Retrocesso para o evento mais próximo;
- 8 – Fazer alteração do fuso-horário da *timeline*;
- 9 – Procurar por eventos.

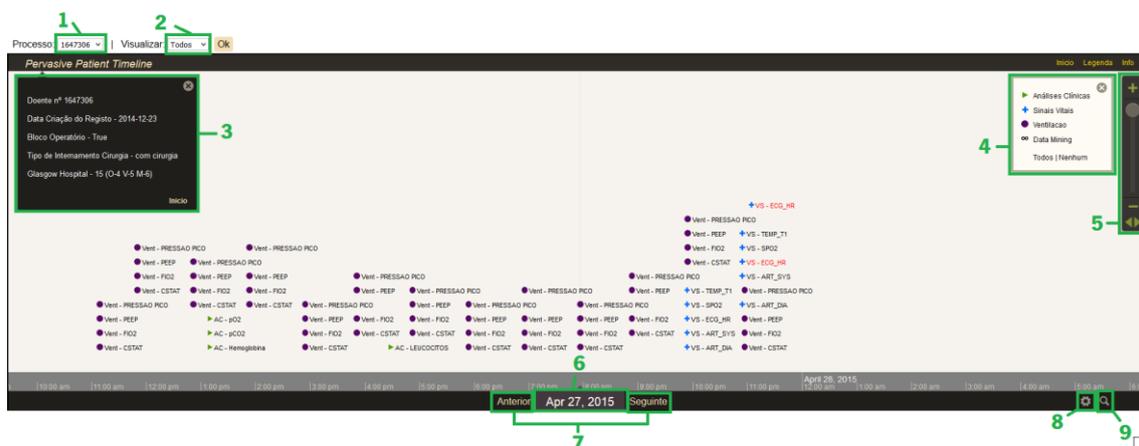


Figura 13 – Representação da *Pervasive Patient Timeline*

Os dados apresentados na Figura 13 são provenientes de doentes internados na UCI do Hospital de Santo António, que pertence ao CHP. Estes dados foram recolhidos de seis fontes de dados: Monitores de Sinais Vitais, Ventiladores, Folha de Enfermagem Eletrónica, Processo Clínico Eletrónico, Laboratórios e Farmácia.

Algumas destas fontes de dados fornecem dados para modelos de *Data Mining* (DM), que visam prever o estado de saúde futuro do doente e sugerir tratamentos, e para o sistema de alertas que avisa os intensivistas para situações em que os doentes monitorizados apresentam valores fora do intervalo de normalidade durante um período de tempo prolongado. A incorporação de modelos

DM na *timeline* adiciona uma nova funcionalidade aos intensivistas, a de poderem antecipar possíveis complicações nos doentes antes de as mesmas acontecerem.

## 4.5 Monitorização

A fase de monitorização da *timeline* servirá o propósito de avaliar o seu desempenho assim que a fase de implementação, que está neste momento a decorrer na UCI do CHP, estiver terminada. A avaliação do desempenho deverá ser feita pelos utilizadores da *timeline*, ou seja, os intensivistas. Nesse sentido, o objetivo desta fase passa por identificar formas de conseguir fazer uma avaliação da futura performance da *timeline*. Para isso foram criados um conjunto de indicadores:

- Tempo de carregamento da *timeline*;
- Tempo de atualização de dados;
- Índice de agrado com o n° de funcionalidades;
- Índice de facilidade do uso do interface;
- Índice de satisfação face à granularidade;
- Índice de satisfação geral dos utilizadores;
- Percentagem de utilizadores da *timeline*;
- Percentagem de utilizadores que reutilizaram a *timeline*;
- Percentagem de doentes em que o diagnóstico melhorou graças a decisões tomadas através do uso da *timeline*;

Com base neste conjunto de indicadores foi elaborado um questionário (ver Anexo I) que tem como objetivo fazer uma avaliação de diferentes aspetos da *timeline*, obtendo assim uma visão mais específica do desempenho. Os resultados obtidos a partir deste questionário deverão ser analisados e ponderados, de forma a perceber a real apreciação dos utilizadores face à solução desenvolvida e se esta cumpre as suas necessidades. Posto isto deverão ser tomadas medidas de correção aos problemas identificados, com o intuito de tornar a solução mais robusta e do agrado dos utilizadores. Aqui o foco é sempre nas pessoas, nos intensivistas, através da melhoria das suas condições de trabalho, e nos doentes, com a prestação de serviços de melhor qualidade e segurança.

## **Capítulo 5 – Descrição do Estudo de Data Mining**

### **5.1 Compreensão do Negócio**

O processo de tomada de decisão tem um papel muito importante na Medicina Intensiva (MI), devido ao grave estado clínico dos doentes internados nas Unidades de Cuidados Intensivos (UCI). Aqui uma decisão mais acertada pode significar o salvamento de vidas, daí a extrema importância. O aumento do número de dados disponíveis sobre os doentes dificulta o processo de decisão na medida em que os intensivistas têm dificuldade em fazer juízos de valor com tantas variáveis em causa.

É com base neste ideal de melhorar o processo de tomada de decisão que foi observada, por intensivistas, a necessidade relativamente à administração de vasopressores.

O estudo diz respeito a um conjunto de doentes com registo de internamento na UCI do Hospital de Santo António, no âmbito do projeto INTCare e pretende ajudar os intensivistas, prevendo a necessidade de doente vir a precisar de tomar vasopressores, usando para o efeito técnicas de *Data Mining* (DM). O estudo pretende trazer vantagens ao nível dos cuidados dados aos doentes na medida em que ajuda os intensivistas a tomar decisões mais precisas e rigorosas, mas também a nível financeiro e da gestão de recursos já que evita o uso de vasopressores, e consequentemente podem ser tomadas medidas prévias no sentido de evitar essa necessidade prevista.

Sendo assim o estudo tem como principal objetivo criar modelos de DM capazes de prever se um doente vai precisar de tomar vasopressores ou não, com base na informação clínica do doente.

No âmbito deste estudo os agentes vasopressores em análise são a adrenalina, noradrenalina e dopamina.

Para o desenvolvimento do estudo, foi utilizada tecnologia Oracle, mais especificamente o *software* Oracle SQL Developer. Esta ferramenta é um ambiente integrado de desenvolvimento que permite o desenvolvimento e gestão de bases de dados Oracle. Todo o processo de compreensão e preparação dos dados foi executado com recurso ao Oracle SQL Developer, recorrendo ainda à sua extensão Oracle Data Miner para a modelação e avaliação dos modelos de DM.

A finalidade dos modelos DM passa também pela sua incorporação na *Pervasive Patient Timeline*, de forma que os intensivistas possam *consultar* as probabilidades futuras de um determinado doente vir a tomar vasopressores. Esta adição faz parte do princípio de centralização de várias fontes de dados numa só plataforma e de ajuda ao processo de tomada de decisão.

## 5.2 Compreensão dos Dados

Os dados utilizados no estudo foram fornecidos pelo Centro Hospitalar do Porto (CHP), e provêm da atividade do subsistema de aquisição de dados incorporado no sistema INTCare. Os dados usados dizem respeito a doentes da Unidade de Cuidados Intensivos (UCI) do Hospital de Santo António e foram recolhidos de cinco fontes de dados:

- Laboratório – Análises Clínicas (AC);
- Monitores de Cabeceira - Sinais Vitais (SV);
- Processo Clínico Eletrónico (PCE);
- Folha de Enfermagem Eletrónica (FEE);
- Farmácia (FRC).

Desta forma, foram reunidos os dados das cinco fontes e incorporados num só conjunto final de dados, após várias transformações. A compreensão dos dados será feita relativamente ao conjunto final de dados, sendo explicado na preparação dos dados o passos realizados até se atingir esse conjunto final.

O espaço temporal do conjunto final de dados vai desde 6 de janeiro de 2015 a 18 de maio de 2015, correspondendo a 1259 linhas de registos sobre 56 doentes distintos. Deste conjunto de dados fazem parte 25 variáveis: Processo, Episódio, Data, Hora SPO2, ECG\_HR, ART\_SYS, TEMP\_T1, PH, Eritrócitos, Potássio, Glicose, Leucócitos, Lactatos, PCO2, Hemoglobina, Idade, Sexo, Proveniência, Tipo\_Internamento, InternamentoCirurgia, Adrenalina, Noradrenalina, Dopamina e VSPGeral, sendo a última a variável a classificar nos modelos.

A seleção destas variáveis foi baseada nos possíveis fatores que originam um conjunto de doenças, nas quais se usam agentes vasopressores para as combater. Por exemplo no caso da adrenalina esta pode ser usada para combater paragens cardíacas, que tem como possíveis causas a falta de oxigénio, níveis anormais de pH no corpo, excesso ou escassez de potássio, baixa temperatura corporal, níveis de glicose elevados ou baixos, entre outras (American Heart Association, 2005). No caso da dopamina esta é comumente usada para tratar hipotensão aguda que é o resultado de baixa pressão arterial. A noradrenalina pode ser usada para tratar sépsis, à qual é associado um aumento do ritmo cardíaco, respiratório, leucócitos e níveis baixos de oxigénio (Bone et al., 1992, Dellinger et al., 2013).

Por forma a perceber o que cada variável representa fez-se a seguinte descrição dos dados:

- Processo** – número de processo do doente internado na UCI;
- Episódio** – número de episódio do doente internado na UCI;
- Data** – data de um registo relativo ao doente;
- Hora** – hora de um registo relativo ao doente;
- SPO2** – representa a concentração de oxigénio no sangue (%);
- ECG\_HR** – representa o ritmo cardíaco do doente (/min);
- ART\_SYS** – representa a pressão sanguínea (mm/Hg);
- TEMP\_T1** – representa temperatura do doente (°C);
- PH** – representa os níveis de pH do doente;
- Eritrócitos** – representa os níveis de eritrócitos do doente;
- Potássio** – representa os níveis de potássio do doente (mmol/L);
- Glicose** – representa os níveis de glicose do doente (mg/dL);
- Leucócitos** – representa os níveis de leucócitos do doente;
- Lactatos** – representa os níveis de lactatos do doente (mmol/L);
- PCO2** – representa a pressão parcial de dióxido de carbono no sangue do doente (mm/Hg);
- Hemoglobina** – representa os níveis de hemoglobina do doente (mg/dL);
- Idade** – idade do doente internado;
- Sexo** – sexo do doente internado;
- Proveniência** – proveniência do doente internado (ex: bloco operatório, outro hospital, ...);
- Tipo\_Internamento** – nível de urgência do internamento (ex: urgente, programado, ...);
- InternamentoCirurgia** – tipo de internamento (ex: com ou sem cirurgia);
- Adrenalina** – representa se foi administrada (1) ou não (0) adrenalina ao doente;
- Noradrenalina** – representa se foi administrada (1) ou não (0) noradrenalina ao doente;
- Dopamina** – representa se foi administrada (1) ou não (0) dopamina ao doente;
- VSPGeral** – representa se foi administrado pelo menos um (1) ou nenhum (0) dos vasopressores.

Na Tabela 6 pode-se observar de um ponto de vista mais técnico as fontes e tabelas a que cada variável do conjunto final pertence:

**Tabela 6 – Variáveis e respetivas tabelas e fontes**

<b>Fonte</b>	<b>Tabela</b>	<b>Variável</b>
SV	Patient_SV	SPO2
		ECG_HR
		ART_SYS
		TEMP_T1
AC	Patient_AC	PH
		Eritrócitos
		Potássio
		Glicose
		Leucócitos
		Lactatos
		PCO2
		Hemoglobina
PCE/FEE	Patient_PID	Idade
		Sexo
	Patient_Admissao	Proveniência
		Tipo_Internamento
		InternamentoCirurgia
FRC	Patient_Vasopressores	Adrenalina
		Noradrenalina
		Dopamina

As variáveis processo, episódio, data e hora não contemplam na Tabela 6 pois são o resultado de igualdades entre várias tabelas. A variável VSPGeral também não aparece listada pois é criada no conjunto final de dados como resultado da combinação das variáveis Adrenalina, Noradrenalina e Dopamina.

Na realização deste estudo foram executadas duas abordagens, uma com os valores dos sinais vitais e análises clínicas (sem classes) e outra com a passagem desses valores a classes (com classes). Assim foram utilizados dois conjuntos de dados, sendo a abordagem com classes uma derivação da abordagem sem classes. Baseado no facto de terem sido executadas duas

abordagens, é de seguida apresentada a exploração dos dados comuns a ambos os conjuntos, com o objetivo de ter uma visão geral sobre os dados. Para isso foram efetuadas análises estatísticas que visam melhorar a compreensão dos dados.

Na Tabela 7 encontra-se representada uma distribuição qualitativa das variáveis comuns às duas abordagens.

**Tabela 7 – Distribuição qualitativa das variáveis comuns às abordagens Sem Classes e Com Classes**

<b>Variável</b>	<b>Valor</b>	<b>Percentagens</b>
Internamento_Cirurgia	Com Cirurgia	69,18%
	Sem Cirurgia	30,82%
Tipo_Internamento	Urgente	78%
	Programado	22%
Sexo	Masculino	64,26%
	Feminino	35,74%
Dopamina	0	100%
	1	0%
Adrenalina	0	100%
	1	0%
Noradrenalina	0	62,19%
	1	37,81%
Proveniencia	Bloco Operatório	54,17%
	Sala Emergência	23,35%
	Enfermaria	14,14%
	Outro Hospital	4,45%
	Outra UCI	2,46%
	Urgência	1,27%
	Unidade Intermédia	0,16%

## Sem Classes

Para a abordagem sem classes nas variáveis de sinais vitais e análises clínicas foi feita uma análise quantitativa dessas variáveis, onde foram calculados diversos valores estatísticos como se pode ver na Tabela 8.

**Tabela 8 – Análise quantitativa das variáveis na abordagem Sem Classes**

Variável	Fonte	Máximo	Média	Mínimo	Desvio Padrão	Variância	Mediana
SPO2	SV	100	95,82	44,59	6,22	38,64	97,41
ECG_HR	SV	187	83,19	39	17,93	321,50	82
ART_SYS	SV	238,42	129,74	3,08	28,75	826,71	128,84
TEMP_T1	SV	39,63	36,48	34	1,03	1,06	36,51
PH	AC	7,80	7,43	6,96	0,09	0,008	7,44
Eritrócitos	AC	5,16	3,24	1,17	0,54	0,29	3,19
Potássio	AC	7,60	3,78	1,60	0,76	0,58	3,70
Glicose	AC	389	143,69	14	48,58	2360,66	135
Leucócitos	AC	53,18	12,11	1,24	7,92	62,76	10,34
Lactatos	AC	16	2,35	0,30	2,34	5,45	1,70
PCO2	AC	172	41,78	13,70	12,27	150,61	40,40
Hemoglobina	AC	20,60	9,81	2	2,03	4,12	9,60
Idade	PCE/FEE	90	62,78	21	15,50	240,29	65

## 5.3 Preparação dos Dados

Na preparação dos dados foi executado um processo Extract, Transform and Loading (ETL). Este processo conduziu à necessidade de efetuar diversas iterações até os dados estarem no estado de qualidade desejado para as duas abordagens. Inicialmente os dados foram retirados de cinco fontes: os sinais vitais foram retirados de Monitores de Cabeceira, os dados de análises clínicas foram retirados do Laboratório, os dados gerais sobre o doente provêm do Processo Clínico Eletrónico e Folha de Enfermagem Eletrónica, e os dados relativos aos vasopressores provêm da Farmácia. Devido ao formato que os dados apresentavam nas quatro tabelas foi necessário fazer algumas transformações até chegar a um conjunto final de dados com tudo agregado por doente

na mesma tabela e adequado às duas abordagens. O conjunto final de dados da abordagem sem classes foi o primeiro a ser preparado, tendo o conjunto com classes sido preparado a partir deste. As transformações efetuadas para o conjunto sem classes encontram-se descritas em quatro etapas: Sinais Vitais, Análises Clínicas, Vasopressores e Conjunto Final. Para o conjunto com classes as transformações encontram-se explicitadas na etapa Novo Conjunto.

### **Sinais Vitais**

**1º** - A tabela inicial de sinais vitais continha 594065 registos sobre 97 doentes distintos relativamente ao ano de 2015. Tendo em conta a existência de vários registos do mesmo sinal vital para a mesma hora, tomou-se a opção de, segundo indicações clínicas, ter apenas um registo por hora por sinal vital escolhido, nomeadamente o valor registado mais perto do início da hora. Com esta alteração ficaram 64528 registos sobre 97 doentes.

**2º** - De seguida foi efetuada uma troca de linhas por colunas de forma a ser possível ter apenas um registo de doente para uma determinada hora, caso contrário haveria vários registos, um para cada variável, relativamente à mesma hora. O número de registos passou para 13036 e originou o aparecimento de 1501 valores nulos.

**3º** - Relativamente aos valores nulos foi atribuído o valor da última hora com registo sobre a variável em questão. Apesar da ligeira melhoria da qualidade dos dados, a integridade e validade da informação recolhida inicialmente não foi adulterada.

### **Análises Clínicas**

**1º** - Na tabela inicial das análises clínicas existiam 194291 registos relativamente a 187 doentes distintos, no ano de 2015. Observando o mesmo problema dos sinais vitais foi efetuada uma transformação para apresentar um registo por hora e por análise clínica escolhida. A alteração diminui o número de registos para 28599 sobre 187 doentes.

**2º** - Foi aqui também efetuada a troca de linhas por colunas diminuindo mais ainda o número de registos para 6285, e originando 6008 valores nulos.

**3º** - Os valores nulos foram tratados da mesma forma dos sinais vitais, ou seja, foram substituídos pelo valor da última hora com registo, mantendo sempre a integridade e validade da informação recolhida.

## Vasopressores

**1º** - O primeiro passo a ser efetuado na tabela com registos farmacêuticos, foi o de isolar os registos em que os medicamentos administrados a doentes não eram Adrenalina, Noradrenalina e Dopamina. Isto traduz-se na passagem de 78080 registos para 1121 registos, sobre 67 doentes.

**2º** - De seguida foi identificada a existência de 1083 registos com data de fim para a administração do vasopressor nula. Estes nulos foram substituídos com a data de saída do doente, registada em uma tabela do Processo Clínico Eletrónico.

**3º** - Destes registos foram eliminados os que tinha data de início da administração inferior ao ano de 2015, pelo que ficaram a restar 101 registos. Com esta transformação deixaram de aparecer registos de Dopamina visto serem do ano de 2014.

## Conjunto Final (agregação de todos os dados)

**1º** - Foi criada a nova tabela para a agregação de todos os dados, tendo sido primeiro carregados os valores das variáveis dos sinais vitais (13036 registos) e análises clínicas (6285 registos) que apresentavam ter Processo, Data e Hora iguais entre eles, resultando esta igualdade num total de 1948 registos.

**2º** - Posteriormente foi feito um *update* aos valores das variáveis idade e sexo, tendo a variável idade sido calculada através da data de nascimento, registada numa das tabelas do Processo Clínico Eletrónico.

**3º** - De seguida foi a vez de adicionar os valores da Proveniência, Tipo\_Internamento e InternamentoCirurgia. Nestas variáveis houve vários registos que ficaram a nulo, pelo que foram removidos.

**4º** - As variáveis Adrenalina, Noradrenalina e Dopamina foram adicionadas, tendo em conta a tabela de vasopressores. Destas três só a Noradrenalina teve valor positivo para a administração, já que foi o único vasopressor com administração para o mesmo doente nas duas tabelas.

**5º** - Foi feito o *update* aos valores da variável VSPGeral, tendo valor “1” caso pelo menos um dos vasopressores tenha sido tomado ou “0” caso nenhum tenha sido administrado.

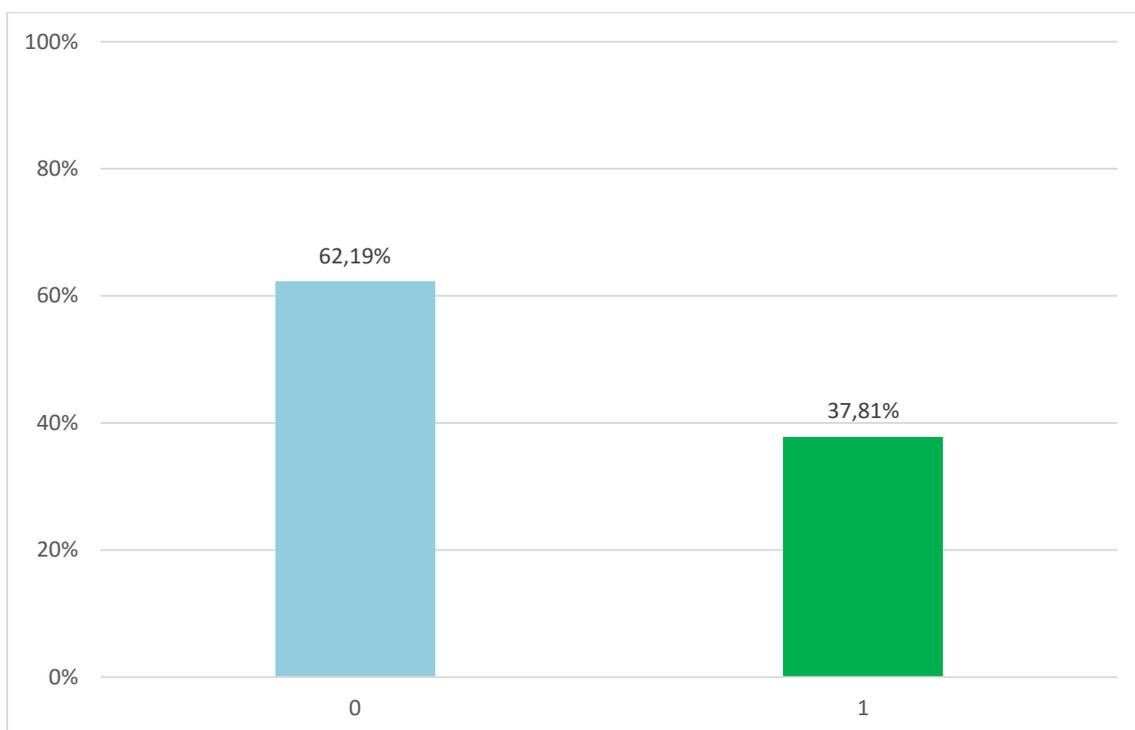
**6º** - Por último foi feita uma análise geral aos dados, e foram tratados os erros nos valores das variáveis das análises clínicas. Os valores completamente desfasados do contexto levaram à eliminação do registo, os valores com erros de escrita foram meramente corrigidos.

A execução de todas estas tarefas deu origem ao conjunto final de dados, tendo para isso sido eliminados 689 registos, que possuíam valores nulos ou com erros, do valor inicial de 1948

registos. Isto resulta num conjunto final de dados com 1259 registos, tal como referenciado na compreensão dos dados.

### Variável VSPGeral

A variável VSPGeral apresenta ter as mesmas percentagens de 37,81% e 62,19%, respetivamente, na administração (1) e não administração (0) de vasopressores que a variável Noradrenalina. Isto deve-se ao facto de as variáveis Adrenalina e Dopamina, não terem sido nunca administradas nos doentes do conjunto de dados, como se pode observar através da comparação entre a Figura 14 e a Tabela 7.



**Figura 14 – Distribuição qualitativa da variável VSPGeral**

### Novo Conjunto (Com classes)

Na segunda abordagem aos dados, onde foram utilizadas classes, foi usada uma função que permite a avaliação de dados do tipo texto e conseqüente conversão para tipo numérico. O uso desta função permitiu a avaliação dos valores das variáveis dos sinais vitais e análises clínicas de modo a atribuir o respetivo valor de classe. As classes foram atribuídas tendo em conta o significado clínico dos valores das variáveis. Esta decisão prende-se com o facto de a utilização de outros métodos, tais como a distribuição normal, poderem atribuir a classe errada a alguns valores.

Desta forma foram agregados nas classes Crítico Baixo, Normal e Crítico Alto. A classe Crítico representam os valores abaixo dos valores de referência, a classe Normal incorpora os valores que se encontra entre os valores de referência e a classe Crítico Alto inclui os valores acima dos valores de referência. Na Tabela 9 está representada as variáveis e os intervalos de valores definidos para cada uma das classes, pelo que a definição destes valores foi realizada com o apoio e indicação de intensivistas da UCI.

**Tabela 9 – Representação das variáveis e valores de referência por classe**

	<b>Variável</b>	<b>Crítico Baixo</b>	<b>Normal</b>	<b>Crítico Alto</b>
<b>Sinais Vitais</b>	TEMP_T1	34-36	36-38	38-45
	ART SYS	0-90	90-180	180-500
	ECG_HR	0-60	60-120	120-250
	SPO2	0-90	90-100	100-500
<b>Análises Clínicas</b>	Hemoglobina	<12,00	12,0 - 18,0	>18,00
	PH	<7,35	7,35 - 7,45	>7,45
	Leucócitos	<4,00	4,00 - 11,00	>11,00
	Potássio	<3,50	3,50 - 5,30	>5,50
	Lactatos	<0,50	0,50 - 2,20	>2,20
	Glicose	<70,00	70,00 - 105,00	>105,00
	PCO2	<32,00	32,00 - 45,00	>45,00
	Eritrócitos	<4,10	4,10 - 5,50	>5,50

Com a formação de um novo conjunto de dados com os dados de análises clínicas e sinais vitais agregados em classes houve a necessidade de realizar uma distribuição qualitativa dessas variáveis. Os resultados podem ser observados na Tabela 10, onde se encontram representadas a distribuições dos valores Crítico Baixo, Normal e Crítico Alto para cada uma das variáveis.

**Tabela 10 – Distribuição qualitativa das variáveis na abordagem Com Classes**

<b>Variável</b>	<b>Crítico Baixo</b>	<b>Normal</b>	<b>Crítico Alto</b>
TEMP_T1	31,22%	61,56%	7,23%
ART_SYS	5,40%	90,15%	4,45%
ECG_HR	7,39%	89,44%	3,18%
SPO2	7,07%	92,93%	0%
Hemoglobina	88,17%	11,52%	0,32%
PH	14,77%	41,46%	43,76%
Leucócitos	9,37%	45,43%	45,19%
Potássio	31,61%	64,89%	3,49%
Lactatos	1,99%	64,89%	33,12%
Glicose	1,51%	15,65%	82,84%
PCO2	11,91%	60,84%	27,24%
Eritrócitos	92,85%	7,15%	0%

Terminadas as tarefas na preparação dos dados passa-se à etapa de modelação, na qual ocorre a criação dos modelos de DM.

## 5.4 Modelação

Nesta fase, o foco é a criação de modelos que através da utilização de técnicas de DM devolvam resultados relativos ao objetivo do estudo. Dada a natureza do objetivo de DM, a escolha de modelação recaiu sobre modelos do tipo classificação.

Posto isto, foi tomada a opção de agrupar as variáveis em quatro grupos para se perceber de que forma cada grupo influenciaria os resultados. Estes grupos foram utilizados para prever a variável VSPGeral. A seguir é possível ver que grupos foram criados e que variáveis pertencem a esses grupos:

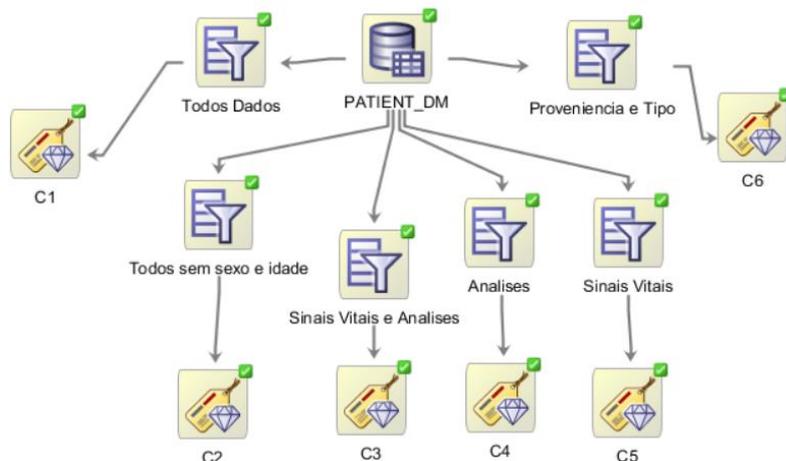
**Sinais Vitais (SV)** = SPO2, ECG\_HR, ART\_SYS, TEMP\_T1

**Análises Clínicas (AC)** = PH, Eritrócitos, Potássio, Glicose, Leucócitos, Lactatos, PCO2, Hemoglobina

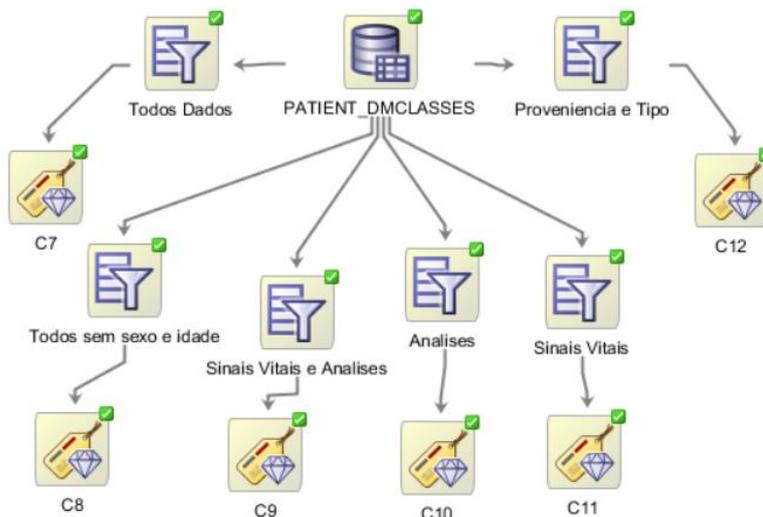
**Admissão Doente (AMD)** = Proveniência, Tipo\_Internamento, InternamentoCirurgia

**Case Mix (CM)** = Idade, Sexo

Com base nos grupos de variáveis formados foram modelados 12 cenários, 6 cenários a utilizar os valores das variáveis (Figura 15) e 6 cenários em que os valores das variáveis dos sinais vitais e análises clínicas foram agregados em classes (Figura 16).



**Figura 15 – Construção dos cenários sem classe**



**Figura 16 – Construção dos modelos com classes**

Todos os cenários criados foram distribuídos utilizando 60% dos dados para treino, enquanto os 40% restantes foram usados para teste. No entanto, nos cenários com classes foram também usados todos os dados para teste, ou seja, *Cross Validation*, por forma a perceber que efeitos poderia ter ao nível dos resultados.

Na Tabela 11, estão representados os cenários modelados e os grupos de variáveis que cada um dos cenários detém.

Tabela 11 – Cenários e grupos de variáveis

Cenários	Grupos
<b>Sem Classe</b>	
C1	SV + AC + AMD + CM
C2	SV + AC + AMD
C3	SV + AC + CM
C4	AC + CM
C5	SV + CM
C6	AMD + CM
<b>Com Classe</b>	
C7	SV + AC + AMD + CM
C8	SV + AC + AMD
C9	SV + AC + CM
C10	AC + CM
C11	SV + CM
C12	AMD + CM

Para cada um dos cenários foram executados quatro algoritmos de DM: *Support Vector Machines* (SVM), Árvore de Decisão (AD), Naïve Bayes (NB) e Modelos Lineares Generalizados (MLG). A combinação dos cenários criados e dos algoritmos utilizados resulta na indução de um total de 72 modelos ([6 cenários s/classe \* 4 algoritmos \* 1 técnica de validação] + [6 cenários c/classe \* 4 algoritmos \* 2 técnicas de validação]). A seguinte expressão representa um modelo geral:

$$M_n = A_x + S_i + T_z + V_t$$

Na expressão  $M_n$  representa um modelo com uma abordagem de classificação ( $A_x$ ), um cenário ( $S_i$ ), um algoritmo ( $T_z$ ) e uma técnica de validação ( $V_t$ ):

$$A_x = \{\text{Classificação}\}$$

$$S_i = \{\text{Cenário 1, ..., Cenário 12}\}$$

$$T_z = \{\text{SVM, NB, MLG, AD}\}$$

$$V_t = \{40\% - 60\%; \text{Cross Validation}\}$$

É também relevante apresentar os parâmetros dos algoritmos usados na construção dos modelos, de forma a perceber os resultados obtidos por cada um dos algoritmos. Na Tabela 12 estão representados os parâmetros utilizados em cada um dos algoritmos.

Tabela 12 – Parâmetros dos algoritmos de DM

<b>Support Vector Machines</b>	
Kernel Function	Gaussian
Tolerance Value	0,001
Active Learning	Yes
<b>Árvores de Decisão</b>	
Homogeneity Metric	Gini
Maximum Depth	7
Minimum Records in a Node	10
Minimum Percent of Records in a Node	0,05
Minimum Records for a Split	20
Minimum Percent of Records for a Split	0,1
<b>Naïve Bayes</b>	
Pairwise and Singleton Threshold	0
<b>Modelos Lineares Generalizados</b>	
Reference Class name	1
Missing Value Treatment	Mean Mode
Ridge Regression	Disable
Approximate Computation	Disable

## 5.5 Avaliação

A fase de avaliação é bastante importante visto ser o momento em que são analisados os resultados e efetuadas as comparações entre os diversos cenários desenvolvidos, para que assim se consiga perceber se os resultados esperados foram atingidos e qual dos cenários apresenta melhor resultado.

A avaliação aos diversos cenários foi realizada recorrendo a curvas *Receiver Operating Characteristic* (ROC) e a matrizes de confusão. Da construção da matriz de confusão foram derivadas três métricas para análise: Sensibilidade, Especificidade e Acuidade. Cada uma destas três métricas procura avaliar diferentes tipos de previsões. Neste caso em específico, se o objetivo for prever se um doente vai precisar que lhe seja administrado um vasopressor então deve ser escolhido o cenário com maior sensibilidade, caso a previsão diga respeito ao doente não vir a

precisar de tomar um vasopressor, deve ser então usado o cenário com maior valor de especificidade. A acuidade por sua vez dá a probabilidade de prever ambas as situações, ou seja, a probabilidade de precisar ou não de tomar vasopressores.

Tendo em conta os resultados esperados foram definidos limites a atingir pelas três métricas para que fossem elegíveis para posterior comparação entre cenários. Para a sensibilidade foi definida uma probabilidade de 85%, para a especificidade uma probabilidade de 60% e para a acuidade 70%.

Posto isto foram efetuadas três análises que podem ser classificadas da seguinte forma:

- Sem Classes (SCL) – diz respeito à abordagem sem classes na qual foi usada uma distribuição 40/60 (%) para teste/treino;
- Com Classes 1 (CL1) – agregou-se dados em classes e foi também usada uma distribuição 40/60 (%) para teste/treino;
- Com Classes 2 (CL2) – usou-se também os dados agregados em classes, mas desta vez foram usados todos os dados para teste/treino, ou seja, *Cross Validation*.

Na Tabela 13 é possível observar os três melhores resultados obtidos para cada uma das três métricas, conforme a análise realizada, e de acordo com os limites impostos para cada métrica.

**Tabela 13 – Três melhores modelos por métrica, e por análise**

<b>SCL</b>			<b>CL1</b>			<b>CL2</b>		
<b>Sensibilidade</b>			<b>Sensibilidade</b>			<b>Sensibilidade</b>		
Cenário	Algoritmo	Valor	Cenário	Algoritmo	Valor	Cenário	Algoritmo	Valor
C1	SVM	90,72%	C6	SVM	94,18%	C6	SVM	94,94%
C1	NB	89,22%	C1	AD	93,18%	C3	SVM	94,44%
C3	NB	88,68%	C3/C4	AD	92,89%	C1	SVM	94,01%
<b>Especificidade</b>			<b>Especificidade</b>			<b>Especificidade</b>		
Cenário	Algoritmo	Valor	Cenário	Algoritmo	Valor	Cenário	Algoritmo	Valor
C3	SVM	82,72%	C1	SVM	85,14%	C1	SVM	87,57%
C4	SVM	76,00%	C3	SVM	79,17%	C4	SVM	85,51%
C1	NB	70,44%	C2	SVM	75,94%	C3	SVM	83,49%
<b>Acuidade</b>			<b>Acuidade</b>			<b>Acuidade</b>		
Cenário	Algoritmo	Valor	Cenário	Algoritmo	Valor	Cenário	Algoritmo	Valor

C3	SVM	86,02%	C1	SVM	89,61%	C1	SVM	91,50%
C4	SVM	82,84%	C3	SVM	87,29%	C4	SVM	90,23%
C1	SVM	81,14%	C6	SVM	84,75%	C3	SVM	89,91%

Numa primeira observação da Tabela 13 podemos ver que os modelos são bastante sensíveis, pelo que vão ao encontro do que mais se esperava, ou seja, modelos capazes de prever a probabilidade de um doente vir a precisar de vasopressores, e que os algoritmos presentes são o SVM, NB e AD, sendo que o SVM é o que está mais presente.

A análise SCL teve melhor resultado ao nível da sensibilidade no cenário C1 e com o algoritmo SVM, obtendo uma probabilidade de acerto de 90,72%. No que toca à especificidade, o melhor resultado figura no cenário C3 e no algoritmo SVM, com uma percentagem de 82,72%. Para a acuidade o melhor resultado foi obtido no cenário C3 e no algoritmo SVM, tal como a especificidade, atingindo um valor de 86,02%

Na análise CL1 o modelo mais sensível é o cenário C6 utilizando o algoritmo SVM tendo aumentado a percentagem para 94,18% (+3,46%). Também na especificidade e acuidade se observa um aumento nos valores do melhor modelo, com um valor de 85,14% (+2,72%) no cenário C1 e algoritmo SVM e uma probabilidade de acerto de 89,61% (+3,59%) no cenário C1 e algoritmo SVM, respetivamente. De uma forma geral é possível assumir que a agregação dos valores dos sinais vitais e análises clínicas em classes se traduziu numa melhoria algo considerável dos resultados de todas as métricas.

No caso da análise CL2 houve também uma melhoria em todas as métricas face à análise CL1, devendo-se isto à aplicação de *Cross Validation*. A métrica sensibilidade sofreu um ligeiro aumento para 94,94% (+0,76%), mantendo como melhor modelo o mesmo que na análise CL1. Já as métricas especificidade e acuidade tiveram um aumento mais significativo para 87,57% (+2,43%) e 91,50% (+1,89%), respetivamente, e em ambas se mantiveram os mesmos modelos como sendo os melhores.

Tendo em conta as três análises pode-se concluir que a agregação dos dados em classes e o uso de *Cross Validation* trouxe melhorias nas capacidades de previsão dos modelos.

No que toca às análises das curvas ROC, a análise SCL atingiu a melhor área sob a curva no cenário C1 e algoritmo NB com uma percentagem de 87,44%. A Figura 17 ilustra a melhor curva ROC da análise SCL.

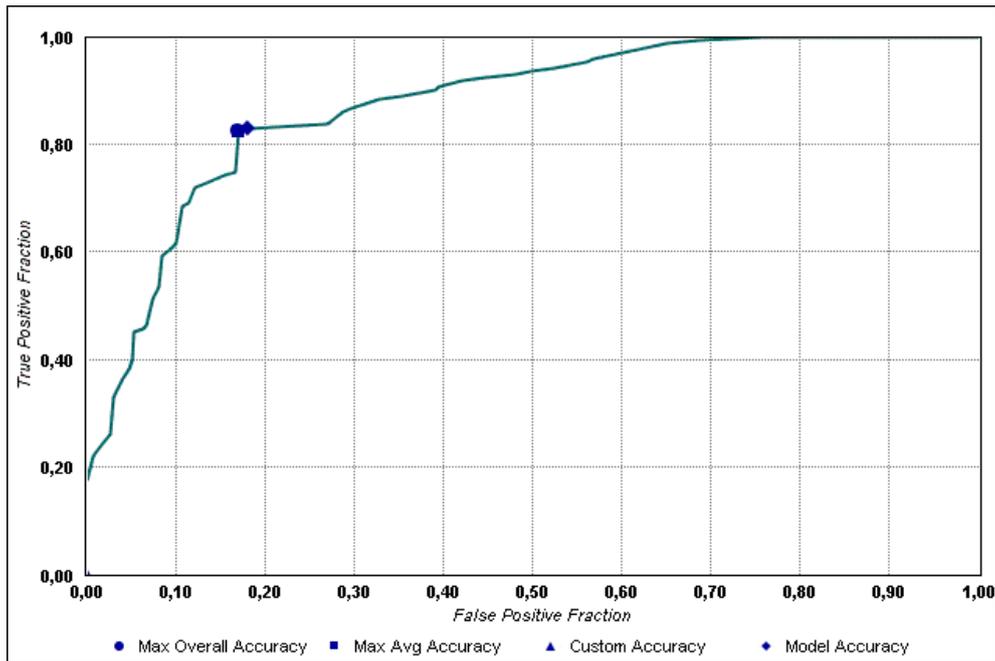


Figura 17 – Melhor curva ROC da análise SCL

Para a análise CL1 a melhor curva ROC é a do modelo com o cenário C7 e algoritmo SVM onde foi verificada uma área sob a curva de 91.36%. Este aumento pode ser atribuído ao facto de os dados dos sinais vitais e análises clínicas terem sido agregados em classes. Na Figura 18 é possível ver representado o gráfico da melhor curva ROC da análise CL1.

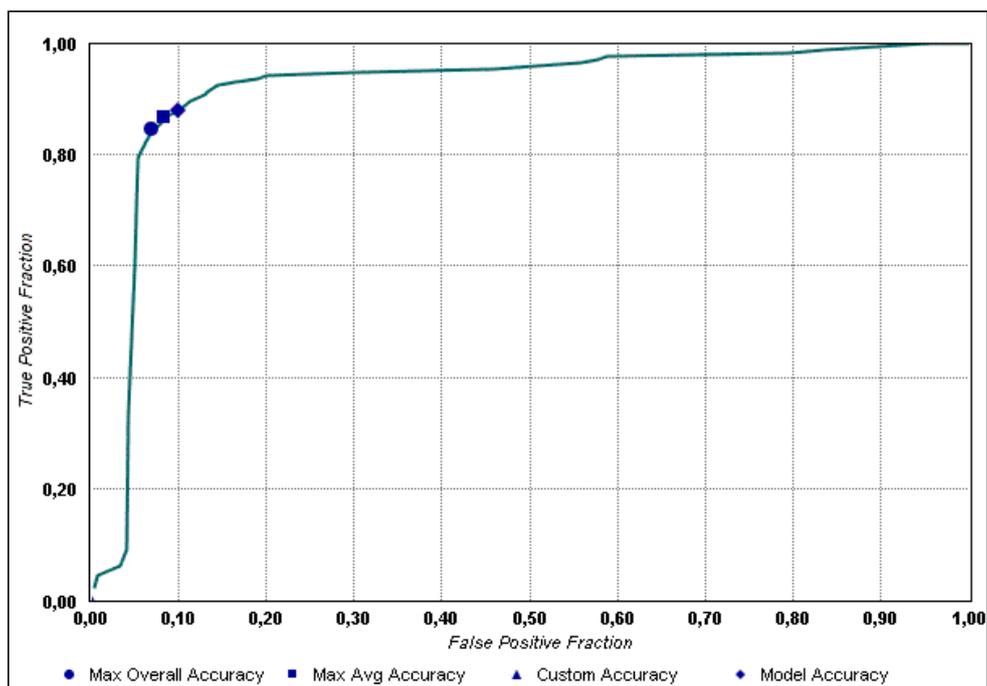
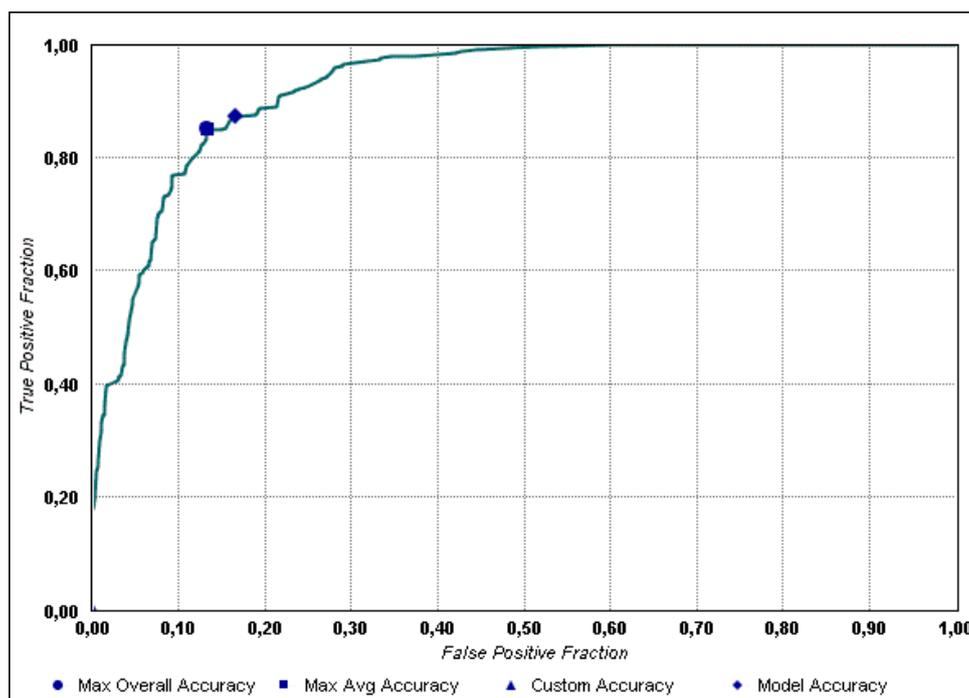


Figura 18 – Melhor curva ROC da análise CL1

No caso da análise CL2, também foi possível verificar um aumento da área sob a curva para 92,77%. Apesar do cenário ser o mesmo que na análise CL1, o algoritmo em que agora se verificou o melhor resultado foi no MLG. O aumento de uma análise para outra pode ser neste caso inculido ao facto de ter sido utilizado *Cross Validation*. A Figura 19 ilustra a melhor curva ROC da análise CL2.

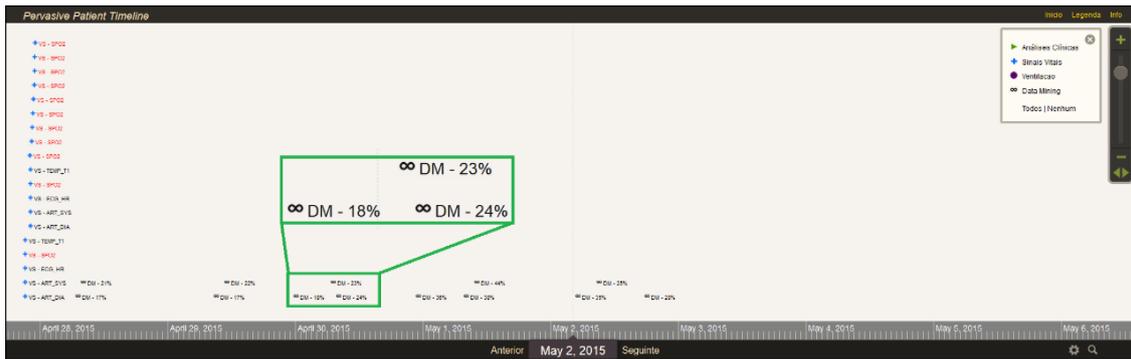


**Figura 19 – Melhor curva ROC da análise CL2**

De forma geral, pode-se concluir da análise às curvas ROC que, numa primeira instância, a agregação de dados em classes e, numa segunda instância, o uso de *Cross Validation* tiveram efeito positivo ao nível dos resultados obtidos, tendo sido observado o melhor resultado na conjugação destas duas alterações.

## 5.6 Implementação

A fase de implementação passa pela adição dos modelos desenvolvidos à *Pervasive Patient Timeline*. Aqui os intensivistas passam a ter a possibilidade de ver, para o doente selecionado, previsões futuras relativamente à probabilidade de tomar vasopressores, servindo como mais um elemento a ajudar ao processo de tomada de decisão nas Unidades de Cuidados Intensivos (UCI). Na Figura 20 está representado um exemplo de um doente no qual contém previsões relativamente à necessidade de tomar vasopressores.



**Figura 20 – Modelos DM na *Pervasive Patient Timeline***

É possível observar que para diferentes momentos são apresentadas previsões sobre a necessidade tomar vasopressores. Isto fornece novo conhecimento aos intensivistas na medida em que lhes apresenta a probabilidade do doente vir a precisar de um vasopressor em determinado momento. Face a isto, os intensivistas podem agir antecipadamente, de forma a tentar assegurar que essa necessidade deixe de existir.



## **Capítulo 6 – Discussão do Trabalho Realizado**

O fim do desenvolvimento de um projeto deve ser seguido de um momento de reflexão, no qual se deve avaliar todo o processo de desenvolvimento, bem como resultado final do projeto, e verificar se os resultados obtidos satisfazem as necessidades observadas inicialmente.

Na definição do objetivo principal do trabalho foi formulada uma questão de investigação ao qual este trabalho deveria responder. Ao longo deste trabalho foram reveladas as formas de como se espera que as *pervasive timelines* possam auxiliar os intensivistas nas suas decisões. O desenvolvimento da *Pervasive Patient Timeline* espelha as capacidades que podem ter as *timelines* e as vantagens que oferecem aos intensivistas. Desta forma, o objetivo de desenvolver uma *pervasive timeline* interativa capaz de providenciar novo conhecimento de forma a facilitar o processo de tomada de decisão considera-se cumprido. As funcionalidades principais da *timeline* foram implementadas, sendo adicionadas outras tendo em conta o contexto em que a mesma foi desenvolvida. A informação clínica do doente, e resultados de sinais vitais, ventilação, análises clínicas e previsões *Data Mining* (DM) foram adicionados à solução, cumprindo o expectável e comprovando a viabilidade da mesma. A observação de todo este conjunto de informação num só local permitirá os intensivistas analisar todos os dados de uma forma intuitiva, fácil e rápida, permitindo-os adquirir novas linhas de pensamento que os levem a considerar relações entre os dados que de outra forma poderiam não considerar.

Relativamente aos modelos de DM induzidos e incorporados na *Pervasive Patient Timeline*, os resultados atingidos são satisfatórios, com resultados acima dos 90%. Este processo sofreu várias transformações, devido à necessidade de colocar os dados num estado de qualidade apropriado para análise. A modelação foi dividida em duas abordagens: uma com dados sem classes e outra com dados em classes. A realização de análises de três tipos (SCL, CL1 e CL2) permitiu tirar conclusões relativamente ao uso dos dados sem classes, com classes e com o uso de *Cross Validation* no teste/treino dos dados. No geral, o resultado dos modelos foram sempre aumentando, primeiro com a agregação em classes e em segundo com o uso de *Cross Validation* e classes, tendo sido atingido o melhor resultado com uma sensibilidade de 94,94% neste último. Desta forma pode-se afirmar que através dos modelos desenvolvidos é possível fazer previsões quanto à futura necessidade de um doente tomar vasopressores.

O resultado deste projeto é assim relevante pois vai facilitar o processo de tomada de decisão com que os intensivistas têm que lidar diariamente e fornecer aos doentes um cuidado de saúde mais preciso.

## **Capítulo 7 – Conclusão**

São vários os constituintes deste trabalho e distinta a forma como contribuíram. A realização de uma revisão de literatura foi valiosa para o trabalho, na medida em que permitiu fazer uma contextualização relativamente às diversas áreas de estudo que se encontravam relacionadas com a *Pervasive Patient Timeline* e a previsão da administração de vasopressores através de modelos de *Data Mining* (DM).

O uso de metodologias como forma de orientar o trabalho realizado foi também uma parte fulcral durante todo o trabalho. Ao todo foram usadas duas metodologias e um modelo: o *Design Science Research* (DSR), o *Cross Industry Process for Data Mining* (CRISP-DM) e o Modelo do Processo de Tomada de Decisão (MPTD). O DSR é a metodologia de investigação que esteve presente durante todo o trabalho realizado, servindo como orientação ao longo do processo de investigação e procurando dar resposta à questão de investigação formulada, tendo sido aplicada em simultâneo com o MPTD e CRISP-DM. O MPTD é um modelo para o desenvolvimento de Sistemas de Apoio à Decisão, pelo que foi seguido para o desenvolvimento da *Pervasive Patient Timeline*. A sua utilização procurou assegurar que a solução desenvolvida era capaz de facilitar o processo de tomada de decisão no qual se foca este modelo. Por sua vez, a metodologia CRISP-DM, sendo um *standard* na área de DM, foi aplicada como forma de facilitar o processo de Descoberta de Conhecimento em Base de Dados, auxiliando, neste caso, na indução de modelos de DM.

A realização deste trabalho culminou com o desenvolvimento de um protótipo da *Pervasive Patient Timeline* que veio demonstrar como a apresentação da informação de uma nova forma pode não só facilitar o trabalho dos intensivistas, mas também melhorar os cuidados prestados aos doentes, através da diminuição dos erros médicos, que é um assunto de extrema importância. O desenvolvimento deste trabalho partiu de uma necessidade observada por intensivistas e visa fazer parte de um processo de repensar a forma de como se pode melhorar a qualidade dos cuidados de saúde prestados nas Unidades de Cuidados Intensivos.

Sendo a adição de previsões futuras recorrendo a modelos de DM algo essencial para a *timeline*, foram desenvolvidos modelos de DM que visam prever a probabilidade de um doente vir a necessitar de tomar vasopressores. O desenvolvimento destes modelos de DM encontra-se numa ótica de reduzir a incerteza das decisões em ambiente clínico em prol dos intensivistas e doentes. O trabalho foi executado no âmbito da fase II do projeto de investigação INTCare, em atividade na UCI do Hospital de Santo António, e resulta da parceria entre a Universidade do Minho e o Centro

Hospitalar do Porto. Os dados usados no trabalho são resultado da atividade do sistema INTCare, que atua em tempo real e de forma automática.

De uma forma geral, o uso da *Pervasive Patient Timeline* irá permitir aos intensivistas aceder a dados clínicos de vários tipos e fontes, tudo numa só plataforma. Assim torna-se mais fácil de aceder aos dados para fazer análises e permite a criação de novo conhecimento através da visualização de possíveis relações entre os dados apresentados na *timeline*, com o intuito de gerar um diagnóstico.

A nível científico foi desenvolvida uma nova solução de visualização de informação que pode ser explorada pela comunidade científica. Com este trabalho ficou provada a viabilidade de adotar as *timelines* como forma de visualizar a informação clínica dos doentes (dados históricos registados em tempo real e previsões futuras). É realçada a importância de uma boa e agradável representação e apresentação dos dados, facto que pode influenciar positivamente as decisões dos intensivistas, bem como a existência de apenas uma fonte que aglomera, homogeneiza e apresenta todo um conjunto de dados distintos, permitindo um acesso mais rápido, fácil e intuitivo. Face a isto os intensivistas podem adquirir novo conhecimento tornando as suas decisões mais fiáveis e precisas.

Por fim, com vista à comunicação do trabalho desenvolvido foram escritos três artigos científicos para revistas e conferências, que podem ser consultados no Anexo II.

## **7.1 Análise dos Riscos**

Tendo em conta o facto de o trabalho ter por base a realização de um projeto, foi desenvolvido um modelo de gestão riscos para acautelar o sucesso do projeto. Na Tabela 14 encontram-se documentados os riscos associados ao trabalho desenvolvido. Em cada um dos riscos é verificado se ocorreu durante o projeto, o impacto que teve, em que numa escala de 1 a 5, o 5 é o nível mais elevado, e qual a seriedade, que é o resultado da multiplicação entre a probabilidade que foi estimada inicialmente e o impacto. Existe também associada a cada risco uma ação atenuante que foi executada sempre que um algum problema ocorreu.

Tabela 14 – Modelo de Análise de Riscos

<b>ID</b>	<b>Descrição</b>	<b>Verificação</b>	<b>Impacto [1 – 5]</b>	<b>Seriedade [1 – 25]</b>	<b>Ação Atenuante (AA)</b>
1	Perda de dados	Não	5	15	-
2	Incumprimento do tempo previsto para as tarefas	Não	4	16	-
3	Complexidade do projeto	Sim	5	20	Dedicação de tempo extra, esclarecimento de dúvidas.
4	Avaria de <i>Hardware</i>	Sim	3	3	Substituição por novo equipamento.
5	Inexperiência no uso da tecnologia	Não	4	16	-



## Referências

- American Heart Association (2005) '7.2: Management of Cardiac Arrest', *Circulation*, 112, pp. 60.
- Bayes, M. and Price, M. (1763) 'An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFRS', *Philosophical Transactions (1683-1775)*, pp. 370-418.
- Berner, E. S. (2009) 'Clinical decision support systems: state of the art', *AHRQ Publication*, (09-0069), pp. 4-26.
- Berner, E. S. and La Lande, T. J. (2007) 'Overview of Clinical Decision Support Systems', in Berner, E.S., FACMI & FHIMSS (eds.) *Clinical Decision Support Systems*. New York: Springer.
- Bone, R. C., Balk, R. A., Cerra, F. B., Dellinger, R. P., Fein, A. M., Knaus, W. A., Schein, R. and Sibbald, W. J. (1992) 'Definitions for sepsis and organ failure and guidelines for the use of innovative therapies in sepsis. The ACCP/SCCM Consensus Conference Committee. American College of Chest Physicians/Society of Critical Care Medicine', *Chest Journal*, 101(6), pp. 1644-1655.
- Braga, P., Portela, F., Santos, M. F. and Rua, F. 'Data Mining Models to Predict Patient's Readmission in Intensive Care Units', *ICAART - International Conference on Agents and Artificial Intelligence*.
- Charette, R. N. (2012) *Troubled HealthSMART System Finally Canceled in Victoria Australia - IEEE Spectrum*. IEEE Spectrum. Available at: <http://spectrum.ieee.org/riskfactor/computing/it/troubled-healthsmart-system-finally-cancelled-in-victoria-australia> (Accessed: February 10 2015).
- Conti, M. a. D. S. K. a. B. C. a. K. M. a. N. L. M. a. P. A. a. R. G. a. T. s. G. a. T. G. a. (2012) 'Looking ahead in pervasive computing: Challenges and opportunities in the era of cyber--physical convergence', *Pervasive and Mobile Computing*, 8(1), pp. 2--21.
- Cook, D. J. and Das, S. K. (2007) 'How smart are our environments? An updated look at the state of the art', *Pervasive and mobile computing*, 3(2), pp. 53--73.
- Cook, D. J. and Das, S. K. (2012) 'Pervasive computing at scale: Transforming the state of the art', *Pervasive and Mobile Computing*, 8(1), pp. 22--35.
- Delen, D., Walker, G. and Kadam, A. (2005) 'Predicting breast cancer survivability: a comparison of three data mining methods', *Artificial intelligence in medicine*, 34(2), pp. 113-127.
- Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., Sevransky, J. E., Sprung, C. L., Douglas, I. S. and Jaeschke, R. (2013) 'Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock, 2012', *Intensive care medicine*, 39(2), pp. 165-228.
- Direção Geral de Saúde, Saúde, D.G.d. (2003) *Cuidados Intensivos: Recomendações para o seu desenvolvimento*. Lisboa.
- Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. (1996) 'From data mining to knowledge discovery in databases', *AI magazine*, 17(3), pp. 37.
- Geiger, K. (1985) 'Application of Computer Systems in Critical Care', *Computers in Critical Care and Pulmonary Medicine*: Springer Berlin Heidelberg, pp. 1-7.
- Han, J., Kamber, M. and Pei, J. (2011) *Data mining: concepts and techniques: concepts and techniques*. Elsevier.
- Hardin, J. M. and Chhieng, D. C. (2007) 'Data Mining and Clinical Decision Support Systems', in Berner, E.S., FACMI & FHIMSS (eds.) *Clinical Decision Support Systems*. New York: Springer.

- Haux, R., Ammenwerth, E., Winter, A. and Brigl, B. (2004) *Strategic information management in hospitals: an introduction to hospital information systems*. Springer, p. 274.
- Health Level Seven International (2007) *About Health Level Seven International*. Available at: <http://www.hl7.org/about/index.cfm?ref=nav> (Accessed: February 16 2015).
- Khalifa, M. (2014) 'Clinical Decision Support: Strategies for Success', *Procedia Computer Science*, 37(0), pp. 422-427.
- Kohn, L. T., Corrigan, J. M. and Donaldson, M. S. (2000) *To Err Is Human:: Building a Safer Health System*. National Academies Press.
- Lewis, F. (2004) 'Smart Environments: Technologies, Protocols, and Applications', *John Wiley, New York*.
- Maimon, O. and Rokach, L. (2005) 'Introduction to knowledge discovery in databases', *Data Mining and Knowledge Discovery Handbook*: Springer, pp. 1-17.
- Marakas, G. M. (2003) *Decision support systems in the 21st century*. Prentice Hall Upper Saddle River, NJ.
- Marcelino, I., Lopes, D., Reis, M., Silva, F., Laza, R. and Pereira, A. (2015) 'Using the eServices platform for detecting behavior patterns deviation in the elderly assisted living: a case study', *BioMed research international*, 2015.
- McDonald, C. and Gardner, R. (1987) 'Computerized Management of Intensive Care Patients', *Images, Signals and Devices M. D. Computing: Benchmark Papers*: Springer New York, pp. 31-45.
- McKenney, J. L. and Scott, M. M. (1984) *Management Decision Systems: Computer-Based Support for Decision Making*. Harvard Business School Press.
- Medscape (2015a) *Dopamine dosing, indications, interactions, adverse effects and more*. Medscape: Medscape. Available at: <http://reference.medscape.com/drug/intropin-dopamine-342435> (Accessed: June 22 2015).
- Medscape (2015b) *Epinephrine dosing, indications, interactions, adverse effects and more*. Medscape: Medscape. Available at: <http://reference.medscape.com/drug/epipen-jr-epinephrine-342437> (Accessed: June 22 2015).
- Medscape (2015c) *Norepinephrine dosing, indications, interactions, adverse effects and more*. Medscape: Medscape. Available at: <http://reference.medscape.com/drug/levarterenol-levophed-norepinephrine-342443#0> (Accessed: June 22 2015).
- Meyfroidt, G., Güiza, F., Ramon, J. and Bruynooghe, M. (2009) 'Machine learning techniques to examine large patient databases', *Best Practice & Research Clinical Anaesthesiology*, 23(1), pp. 127-143.
- Mitchell, W. J. (1980) 'Spatial form in literature: toward a general theory', *Critical Inquiry*, pp. 539-567.
- Morik, K., Brockhausen, P. and Joachims, T. (1999) *Combining statistical learning with a knowledge-based approach: a case study in intensive care monitoring*: Technical Report, SFB 475: Komplexitätsreduktion in Multivariaten Datenstrukturen, Universität Dortmund.
- Moro, S., Laureano, R. and Cortez, P. 'Using data mining for bank direct marketing: An application of the crisp-dm methodology'. *Proceedings of European Simulation and Modelling Conference-ESM'2011*: Eurosis, 117-121.
- Morris, A. and Gardner, A. (1992) 'Computer Applications', in Hall, J., Schmidt, G. & Wood, L. (eds.) *Principles of Critical Care*. New York: McGraw-Hill, pp. 500-514.
- Musen, M. A., Shahar, Y. and Shortliffe, E. H. (2006) *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. Health Informatics New York: Springer New York, p. 701-707.
- Oliveira, J. (2001) 'A shotgun wedding: business decision support meets clinical decision support', *Journal of healthcare information management: JHIM*, 16(4), pp. 28-33.

- Oliveira, S., Portela, F. and Santos, M. F. (2014) 'Pervasive Universal Gateway for Medical Devices.', *Recent Advances in Electrical Engineering and Education Technologies (SCI 2014)*, pp. 205-210.
- Overgaard, C. B. and Džavík, V. (2008) 'Inotropes and vasopressors review of physiology and clinical use in cardiovascular disease', *Circulation*, 118(10), pp. 1047-1056.
- Oztekin, A., Delen, D. and Kong, Z. J. (2009) 'Predicting the graft survival for heart–lung transplantation patients: An integrated data mining methodology', *international journal of medical informatics*, 78(12), pp. e84-e96.
- Paiva, R., Bittencourt, I. I., Pacheco, H., da Silva, A. P., Jacques, P. and Isotani, S. (2012) 'Mineração de Dados e a Gestão Inteligente da Aprendizagem: Desafios e Direcionamentos'.
- Peffer, K., Tuunanen, T., Rothenberger, M. A. and Chatterjee, S. (2007) 'A design science research methodology for information systems research', *Journal of management information systems*, 24(3), pp. 45-77.
- Portela, C. F., Santos, M. F., Silva, Á., Machado, J. and Abelha, A. (2011) 'Enabling a Pervasive Approach for Intelligent Decision Support in Critical Health Care', in Cruz-Cunha, M., Varajão, J., Powell, P. & Martinho, R. (eds.) *ENTERprise Information Systems Communications in Computer and Information Science*: Springer Berlin Heidelberg, pp. 233-243.
- Portela, F., Santos, M., Machado, J., Abelha, A., Silva, Á. and Rua, F. (2014a) 'Pervasive and Intelligent Decision Support in Intensive Medicine – The Complete Picture', *Information Technology in Bio- and Medical Informatics Lecture Notes in Computer Science*: Springer International Publishing, pp. 87-102.
- Portela, F., Veloso, R., Santos, M. F., Machado, J. M., Abelha, A., Silva, Á., Rua, F. and Oliveira, S. M. C. 2014b. Predict hourly patient discharge probability in intensive care units using data mining.
- Ramon, J., Fierens, D., Güiza, F., Meyfroidt, G., Blockeel, H., Bruynooghe, M. and Van Den Berghe, G. (2007) 'Mining data from intensive care patients', *Advanced Engineering Informatics*, 21(3), pp. 243-256.
- Richardson, M., Reilly, W. and Kuntz, J. (2008) *How Timeglider Works*. Available at: [http://timeglider.com/how\\_it\\_works](http://timeglider.com/how_it_works) (Accessed: December 8 2014).
- Rosenberg, D. and Grafton, A. (2010) *Cartographies of time: A history of the timeline*. Princeton Architectural Press.
- SA Health *Enterprise Patient Administration System (EPAS) :: SA Health*. Available at: <http://www.sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+internet/health+reform/ehealth/enterprise+patient+administration+system+epas> (Accessed: February 10 2015).
- SA Health *Implementation of EPAS :: SA Health*. Available at: <http://www.sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+internet/health+reform/ehealth/enterprise+patient+administration+system+epas/implementation+of+epas> (Accessed: February 10 2015).
- Salzhauer, A. (2005) 'Is there a patient in the house', *harvard business review*, 32.
- Santos, M. F. and Azevedo, C. S. (2005) *Data mining: descoberta de conhecimento em bases de dados*. FCA editores.
- Santos, M. F., Cortez, P., Gago, P., Silva, Á. and Rua, F. 'Intelligent decision support in Intensive Care Medicine'. *2nd International Conference on Knowledge Engineering and Decision Support, Lisbon, Portugal*, 401-405.
- Santos, M. F., Mathew, W. and Portela, C. F. (2011a) 'Grid Data Mining for Outcome Prediction in Intensive Care Medicine', *ENTERprise Information Systems*, pp. 244-253.
- Santos, M. F., Portela, F. and Vilas-Boas, M. (2011b) 'Intcare: multi-agent approach for real-time intelligent decision support in intensive medicine'.

- Santos, M. F., Portela, F., Vilas-Boas, M., Silva, Á. and Rua, F. (2010) 'Hourly prediction of organ failure and outcome in intensive care based on data mining techniques'.
- Schneider, J. (1997) *Cross Validation*: Carnegie Mellon University - School of Computer Science. Available at: <http://www.cs.cmu.edu/~schneide/tut5/node42.html> (Accessed: October 24 2015).
- Shortliffe, E. H., Axline, S. G., Buchanan, B. G., Merigan, T. C. and Cohen, S. N. (1973) 'An artificial intelligence program to advise physicians regarding antimicrobial therapy', *Computers and Biomedical Research*, 6(6), pp. 544-560.
- Silva, Á. J. B. M. d. (2007) 'Modelos de inteligência artificial na análise da monitorização de eventos clínicos adversos, disfunção/falência de órgãos e prognóstico do doente crítico'.
- SIMILE Project *SIMILE - Web Widget for Visualizing Temporal Data*. Available at: <http://www.simile-widgets.org/timeline/> (Accessed: September 25 2015).
- Simon, H. 1960. *The New Science of Management Decision*. New York: Harper and Row.
- Soni, J., Ansari, U., Sharma, D. and Soni, S. (2011) 'Predictive data mining for medical diagnosis: An overview of heart disease prediction', *International Journal of Computer Applications*, 17(8), pp. 43-48.
- Sprague, J. and Ralph, H. (1980) 'A Framework for the Development of Decision Support Systems', *MIS quarterly*, 4(4), pp. 1-26.
- Suter, P., Armaganidis, A., Beaufils, F., Bonfill, X., Burchardi, H., Cook, D., Fagot-Largeault, A., Thijs, L., Vesconi, S., Williams, A., Le Gall, J. R. and Chang, R. (1994) 'Predicting outcome in ICU patients', *Intensive Care Medicine*, 20(5), pp. 390-397.
- Tape, T. G. (2006) *Interpreting diagnostic tests*: University of Nebraska Medical Center. Available at: <http://gim.unmc.edu/dxtests> (Accessed: September 15 2015).
- TIME Staff (2013) *Nelson Mandela's Extraordinary Life: An Interactive Timeline*. Available at: <http://world.time.com/2013/12/05/nelson-mandelas-extraordinary-life-an-interactive-timeline/> (Accessed: February 11 2015).
- Timeglider JS *Timglider jQuery Plugin/Widget* (Accessed: September 25 2015).
- Tripathi, K. (2011) 'Decision support system is a Tool for making better decisions in the organization', *Indian Journal of Computer Science and Engineering (IJCSSE)*, 2(1), pp. 112-117.
- Trivedi, M., Kern, J., Marcee, A., Grannemann, B., Kleiber, B., Bettinger, T., Altshuler, K. and McClelland, A. (2002) 'Development and implementation of computerized clinical guidelines: barriers and solutions', *Methods of information in medicine*, 41(5), pp. 435-442.
- Turban, E., Sharda, R. and Delen, D. (2014) *Business Intelligence and Analytics: Systems for Decision Support*. 10th edn.: Prentice Hall.
- Vaishnavi, V. and Kuechler, W. (2004) 'Design research in information systems'.
- Varshney, U. (2009) 'Pervasive Computing', *Pervasive Healthcare Computing*: Springer US.
- Veloso, R., Portela, F., Santos, M., Machado, J. M., Abelha, A., Silva, Á. and Rua, F. 'Real-time data mining models for predicting length of stay in intensive care units'. *KMIS 2014- International Conference on Knowledge Management and Information Sharing*.
- Vincent, G. K. and Velkoff, V. A. (2010) *The next four decades: The older population in the United States: 2010 to 2050*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau.
- Vis.js *vis.js - A dynamic, browser based visualization library*. Available at: <http://visjs.org/> (Accessed: September 25 2015).
- Weiser, M. (1991) 'The Computer for the 21st Century', *Scientific American*, (no. 3), pp. 94-104.
- Wickramasinghe, N., Troshani, I., Hill, S. R., Hague, W. and Goldberg, S. (2013) 'A transaction cost assessment of a pervasive technology solution for gestational diabetes', *Healthcare Information Technology Innovation and Sustainability: Frontiers and Adoption: Frontiers and Adoption*, pp. 109.

- Wilson, J. M. (2003) 'Gantt charts: A centenary appreciation', *European Journal of Operational Research*, 149(2), pp. 430-437.
- XAMPP (2015) *About the XAMPP project*. Available at:  
<https://www.apachefriends.org/about.html> (Accessed: October 24 2015).
- Yegnanarayana, B. (2009) *Artificial neural networks*. PHI Learning Pvt. Ltd.
- Zwicker, M., Seitz, J. and Wickramasinghe, N. (2011) 'The telematics infrastructure: The backbone of the German e-health card', *International Journal of Healthcare Delivery Reform Initiatives (IJHDRI)*, 3(4), pp. 16-26.

## Anexos

### Anexo I – Avaliação do Desempenho da *Pervasive Patient Timeline*

Já usou a Pervasive Patient Timeline? \*

- Sim
- Não

Já utilizou mais que uma vez a Pervasive Patient Timeline? \*

- Sim
- Não

Quantos doentes diagnosticou com o uso da timeline? \*

Quantos doentes melhoraram o diagnóstico graças ao uso da timeline? \*

Qual o tempo do carregamento inicial da timeline? \*

- < 3 segundos
- > 3 e < 6 segundos
- > 6 segundos

Qual o tempo de atualização dos dados? \*

(ex: troca de doente)

- < 3 segundos
- > 3 e < 6 segundos
- > 6 segundos

Figura 21 – Avaliação de Desempenho da *Pervasive Patient Timeline* – Parte 1

**Qual o nível de satisfação com a quantidade de informações hospitalares consideradas? \***

- 0 - Nada satisfeito
- 1 - Pouco satisfeito
- 2 - Satisfeito
- 3 - Muito Satisfeito
- 4 - Extremamente Satisfeito

**Qual o nível de satisfação relativamente à granularidade? \***

- 0 - Nada satisfeito
- 1 - Pouco satisfeito
- 2 - Satisfeito
- 3 - Muito Satisfeito
- 4 - Extremamente Satisfeito

**Qual o nível de satisfação com o nº de funcionalidades? \***

- 0 - Nada satisfeito
- 1 - Pouco satisfeito
- 2 - Satisfeito
- 3 - Muito Satisfeito
- 4 - Extremamente Satisfeito

**Qual o nível de facilidade de uso do interface? \***

- 0 - Nada fácil
- 1 - Pouco fácil
- 2 - Fácil
- 3 - Muito fácil
- 4 - Extremamente fácil

**Qual o nível de satisfação geral com a timeline? \***

- 0 - Nada satisfeito
- 1 - Pouco satisfeito
- 2 - Satisfeito
- 3 - Muito Satisfeito
- 4 - Extremamente Satisfeito

**Figura 22 – Avaliação de Desempenho da *Pervasive Patient Timeline* – Parte 2**

## **Anexo II – Artigos Científicos**

### **1. Step Towards a Patient Timeline in Intensive Care Units**

Autores: André Braga, Filipe Portela, Manuel Filipe Santos, José Machado, António Abelha, Álvaro Silva e Fernando Rua

Revista/Conferência: Procedia Computer Science/Health and Social Care Information Systems and Technologies 2015

Resumo: In Intensive Medicine, the presentation of medical information is done in many ways, depending on the type of data collected and stored. The way in which the information is presented can make it difficult for intensivists to quickly understand the patient's condition. When there is the need to cross between several types of clinical data sources the situation is even worse. This research seeks to explore a new way of presenting information about patients, based on the timeframe in which events occur. By developing an interactive Patient Timeline, intensivists will have access to a new environment in real-time where they can consult the patient clinical history and the data collected until the moment. The medical history will be available from the moment in which patients is admitted in the ICU until discharge, allowing intensivist to examine data regarding vital signs, medication, exams, among others. This timeline also intends to, through the use of information and models produced by the INTCare system, combine several clinical data in order to help diagnose the future patients' conditions. This platform will help intensivists to make more accurate decision. This paper presents the first approach of the solution designed.

Palavras-Chave: Intensive Medicine, Intensive Care Units, Patient-centered, Patient Timeline, INTCare

## **2. Data Mining to predict the use of Vasopressors in Intensive Medicine Patients**

Autores: André Braga, Filipe Portela, Manuel Filipe Santos, António Abelha, José Machado, Álvaro Silva e Fernando Rua

Revista/Conferência: Jurnal Teknologli UTM/International Symposium on Health and Medical Sciences 2015

Resumo: The role that decision making process plays in Intensive Medicine is very critical essential due to the bad health condition of the patients that go to Intensive Care Units (ICU) and the need of a quick and accurate decisions. Therefore each decision is crucial, because it can help saving endangered lives. The decision should be always taken in the patient best interest after analyzing all the data available. In the eyes of the intensivists, the ever growing amount of available data concerning the patients, makes it each time more difficult for them to make a decision based on so many information. It is based on this ideal of improving the decision making process, that this work arises and Data Mining models were induced to predict if a patient will need to take a vasopressor, more specifically: Dopamine, Adrenaline or Noradrenaline. This work used real data provided by an Intensive Care Unit and collected in real-time. The data mining model were induced using data from vital sign monitors, laboratory analysis and information about the patient's Electronic Health Record. This study was based in clinical evidences and provided very useful results with a sensitivity around 90%. These models will reduce the need of vasopressor drugs by helping intensivists to act and take accurate decision before the vasopressor be need by the patient. It will improve the patient condition because when the time comes the predicted necessity of the vasopressor will cease to exist due to the early care provided by the intensivist. The decisions can be for example change the therapeutic plan. Overall, the decision making process becomes more reliable and effective and the quality of care given to patients is better.

Palavras-Chave: Intensive Medicine; Intensive Care Units; Patient-centered; Data Mining; INTCare; Vasopressor; Decision Making

### **3. Real-Time models to predict the use of vasopressors in monitored patients**

Autores: André Braga, Filipe Portela, Manuel Filipe Santos, António Abelha, José Machado, Álvaro Silva e Fernando Rua

Revista/Conferência: Lecture Notes in Computer Science/International Conference for Smart Health 2015

Resumo: The needs of reducing human error has been growing in every field of study, and medicine is one of those. Through the implementation of technologies is possible to help in the decision making process of clinics, therefore to reduce the difficulties that are typically faced. This study focuses on easing some of those difficulties by presenting real-time data mining models capable of predicting if a monitored patient, typically admitted in intensive care, will need to take vasopressors. Data Mining models were induced using clinical variables such as vital signs, laboratory analysis, among others. The best model presented a sensitivity of 94.94%. With this model it is possible reducing the misuse of vasopressors acting as prevention. At same time it is offered a better care to patients by anticipating their treatment with vasopressors.

Palavras-Chave: Vasopressors, INTCare, Intensive Medicine, Real-Time, Data Mining, Vital Signs, Laboratory Results.