

数学学院案例分析报告

案例名称：心电案例分析

报告参与人：魏丹怡（202032164）

完成时间：2021.06.09

目 录

| | |
|------------------|----|
| 一、绪论..... | 1 |
| 1、研究背景..... | 1 |
| 2、研究意义..... | 1 |
| 二、数据来源和基本情况..... | 1 |
| 1、数据来源..... | 1 |
| 2、基本情况..... | 1 |
| 三、数据预处理..... | 2 |
| 1、滤波..... | 2 |
| 2、R 波检测..... | 4 |
| 3、特征提取..... | 5 |
| 四、心电图分类..... | 8 |
| 1、SVM 算法..... | 8 |
| 2、评估方法..... | 9 |
| 3、分类结果及对比..... | 9 |
| 五、实践总结..... | 12 |

一、 绪论

1、研究背景

在现代社会，心血管疾病已经成为了近些年最严重的死亡疾病。根据权威机构的研究数据表明，全世界总死亡人数中大约有三分之一的人死于心血管疾病，相当多的病人死亡的原因是没能及时发现病情从而延误治疗。特别是我国，心脏病死亡人数占总死亡人数的 44%，而且随着人民生活水平的不断提高，心脏病的发病率和致死率呈现逐年上升态势。因而，世界医学界一直都重视心脏病的诊断、治疗。医学界通过对信号的特点和规律的探索，可对一些疾病做出早期预测监护和诊断。随着心电生理学研究的开展和深入，心电图在临床心血管疾病的诊断中起到极大作用。

心脏在收缩跳动过程，心肌内的电流产生在前，心肌机械性收缩在后，应用心电图机以图形形式记录这些心肌电流的变化就叫做心电图。正常人的心电图，一个心动周期由 P 波、QRS 波群和 T 波组成。P 波代表心房除极，QRS 波群代表心室除极，T 波代表心室复极。对照正常的波形和间期，就能推论出生理异常的病症。

2、研究意义

二十世纪七八十年肩带，科技日新月异，能够用于临床的心电设备不断推陈出新，各种检测和结果表示的精确度都能达到相当高的水平。伴随着人工智能和计算机科学与技术不断发展，它们的高效性已逐渐得到生物医学工程方向的研究人员的普遍认同，并称为生物医学尖端领域发展的最重要工具之一。生物医学工程研究人员开始把注意力投向这个新兴领域，即研究如何利用计算机自动实现心电图的检测和诊断。医疗诊断与人工智能技术相结合，不仅有效的降低误诊率和医疗诊断费用，而且可以减轻临床医师的诊断难度。因此，这项课题有着非常广阔的前景，值得深入研究。

二、 数据来源和基本情况

1、数据来源

在 1989 年 10 月至 1996 年 6 月的近七年中，使用 Schiller AG 的设备收集了 PTB-XL ECG 数据集的基础波形数据。通过从 Schiller AG 收购原始数据库，将全部使用权转让给了 Schiller AG。PTB。在 Physikalisch-Technische Bundesanstalt (PTB) 的一个长期项目中，整理了这些记录并将其转换为结构化数据库。该数据库已在许多出版物中使用，但是直到现在，访问仍然受到限制。机构伦理委员会批准了匿名数据在开放访问数据库(PTB-2020-1)中的发布。在 2019 年的公开发布过程中，对现有数据库进行了简化，特别是针对机器学习社区的可用性和可访问性。

2、基本情况

所给数据集是从 PTB-XL 心电数据集中选取了两类样本，分别是正常心电数据和心肌梗

塞心电数据，每类样本数为 100。每个样本记录临床单导联的心电图（ECG），记录时间为 10 秒，采样率为 100Hz，即单个样本数据大小为 1000（采样点数）。将数据集以 9：1 划分成训练集和测试集。

表 1 心电数据描述

| 样本数 | 类别 | 描述 |
|-----|---------|-----------------------|
| 100 | NORM（0） | Normal ECG |
| 100 | MI（1） | Myocardial Infarction |

心电数据测试集和训练集中又包含两个 excel 文件，其中一个 是 20 个样本，另一个是 每个样本对应的标签，其中为 0（正常心电）或 1（心肌梗塞心电），样本数据如下图所示。

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U |
|----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|-----------|----------|----------|----------|----------|----------|---|
| 1 | -0.05097 | -0.01792 | 0.05411 | 0.047385 | 0.05698 | 0.003166 | -0.09721 | -0.0698 | 0.000361 | 0.022949 | -0.08462 | -0.04172 | -0.19195 | -0.02112 | -0.06601 | 0.049635 | -0.11029 | -0.05483 | 0.078657 | 0.016789 | |
| 2 | -0.06786 | -0.02847 | 0.073541 | 0.060028 | 0.07606 | 0.008222 | -0.13106 | -0.08966 | 0.007426 | 0.031463 | -0.11085 | -0.05651 | -0.2512 | -0.0281 | -0.08724 | 0.054117 | -0.12379 | -0.07443 | 0.10261 | 0.021221 | |
| 3 | -0.08331 | -0.03854 | 0.098011 | 0.064938 | 0.0949 | 0.01339 | -0.16508 | -0.10968 | 0.014386 | 0.041128 | -0.13567 | -0.07307 | -0.30905 | -0.03507 | -0.10811 | 0.070965 | -0.15096 | -0.07205 | 0.125521 | 0.024566 | |
| 4 | -0.08308 | -0.03181 | 0.099772 | 0.052248 | 0.095289 | 0.023612 | -0.16929 | -0.1001 | 0.023808 | 0.044464 | -0.13092 | -0.0711 | -0.30111 | -0.03554 | -0.10728 | 0.085346 | -0.15321 | 0.089967 | 0.120471 | 0.021629 | |
| 5 | -0.07892 | -0.02244 | 0.106458 | 0.039767 | 0.095922 | 0.034389 | -0.17072 | -0.08491 | 0.029387 | 0.048145 | -0.1287 | -0.0958 | -0.29572 | -0.03345 | -0.10512 | 0.193199 | -0.13976 | 0.279189 | 0.114936 | 0.020179 | |
| 6 | -0.07428 | -0.02855 | 0.107896 | 0.030427 | 0.097268 | 0.041802 | -0.17588 | -0.07048 | 0.030005 | 0.056258 | -0.12889 | -0.16582 | -0.29585 | -0.02916 | -0.10247 | 0.377465 | -0.13107 | 0.310421 | 0.108255 | 0.021821 | |
| 7 | -0.07434 | -0.02823 | 0.115365 | 0.02158 | 0.097939 | 0.047603 | -0.18173 | -0.05592 | 0.029802 | 0.067548 | -0.12871 | -0.17559 | -0.29637 | -0.02283 | -0.09993 | 0.476236 | -0.13252 | 0.307274 | 0.102758 | 0.024382 | |
| 8 | -0.07505 | -0.02557 | 0.149647 | 0.01181 | 0.099255 | 0.057404 | -0.18706 | -0.04168 | 0.034405 | 0.078846 | -0.12968 | -0.13847 | -0.29981 | -0.01338 | -0.09686 | 0.469774 | -0.11423 | 0.302024 | 0.099235 | 0.028018 | |
| 9 | -0.07295 | -0.04027 | 0.157469 | 0.005008 | 0.099548 | 0.065586 | -0.19174 | -0.03294 | 0.038787 | 0.089959 | -0.13072 | -0.13416 | -0.30581 | -0.01411 | -0.09275 | 0.449578 | -0.08539 | 0.149071 | 0.091762 | 0.032784 | |
| 10 | -0.07485 | -0.07976 | 0.136337 | 0.006396 | 0.101253 | 0.070817 | -0.19365 | -0.03126 | 0.04302 | 0.100835 | -0.12926 | -0.11096 | -0.31497 | 0.018222 | -0.09025 | 0.340225 | -0.06481 | -0.02282 | 0.081233 | 0.03661 | |
| 11 | -0.07494 | -0.13025 | 0.126088 | -0.0662 | 0.106557 | 0.079591 | -0.19075 | -0.03175 | 0.048577 | 0.107961 | -0.12614 | -0.04681 | -0.32484 | 0.037367 | -0.08916 | 0.150752 | -0.04488 | -0.0352 | 0.069368 | 0.037989 | |
| 12 | -0.07186 | -0.12531 | 0.110472 | -0.19795 | 0.112837 | 0.0876 | -0.18564 | -0.03505 | 0.050016 | 0.112333 | -0.12365 | -0.03777 | -0.33755 | 0.055641 | -0.08824 | 0.056532 | -0.04082 | -0.01229 | 0.053018 | 0.040015 | |
| 13 | -0.07435 | -0.2149 | 0.072105 | -0.2936 | 0.113754 | 0.086986 | -0.18057 | -0.03895 | 0.050389 | 0.11317 | -0.12095 | -0.07097 | -0.35107 | 0.075154 | -0.08737 | 0.058046 | -0.03818 | -0.00967 | 0.032584 | 0.043386 | |
| 14 | -0.07764 | -0.23 | 0.060593 | -0.38946 | 0.110843 | 0.080011 | -0.17342 | -0.04045 | 0.053413 | 0.114626 | -0.11824 | -0.07502 | -0.36212 | 0.086228 | -0.08803 | 0.058573 | -0.03602 | -0.00215 | 0.013818 | 0.046814 | |
| 15 | -0.07917 | -0.19577 | 0.074577 | -0.44939 | 0.101742 | 0.070585 | -0.16097 | -0.04397 | 0.060195 | 0.114774 | -0.11516 | -0.06997 | -0.36969 | 0.08596 | -0.08857 | 0.0562 | -0.03539 | -0.00225 | -0.00322 | 0.049986 | |
| 16 | -0.08223 | -0.1544 | 0.081034 | -0.39044 | 0.082646 | 0.056606 | -0.14141 | -0.05103 | 0.071409 | 0.114023 | -0.11344 | -0.05862 | -0.37489 | 0.080238 | -0.08683 | 0.054228 | -0.03541 | 0.001066 | -0.01718 | 0.054418 | |
| 17 | -0.08377 | -0.11317 | 0.083545 | -0.27478 | 0.070386 | 0.041199 | -0.11759 | -0.05596 | 0.085996 | 0.112979 | -0.11296 | -0.04619 | -0.37711 | 0.069591 | -0.08438 | 0.055954 | -0.03519 | 0.001315 | -0.023 | 0.058078 | |
| 18 | -0.08123 | -0.07081 | 0.080461 | -0.18046 | 0.07162 | 0.028403 | -0.09701 | -0.05923 | 0.102003 | 0.112019 | -0.11283 | -0.03519 | -0.37707 | 0.055447 | -0.08102 | 0.057062 | -0.03521 | -0.0001 | -0.01678 | 0.058212 | |
| 19 | -0.07838 | -0.03844 | 0.082748 | -0.09031 | 0.078389 | 0.0176 | -0.0808 | -0.06377 | 0.116199 | 0.110982 | -0.11337 | -0.02343 | -0.37473 | 0.043962 | -0.07599 | 0.057921 | -0.03499 | 0.003099 | 0.001195 | 0.055993 | |
| 20 | -0.07621 | -0.03579 | 0.081433 | -0.04003 | 0.086003 | 0.008601 | -0.07001 | -0.06542 | 0.130188 | 0.110217 | -0.11463 | -0.01657 | -0.37306 | 0.035637 | -0.06921 | 0.057091 | -0.03661 | 0.009913 | 0.029217 | 0.052599 | |
| 21 | -0.07399 | -0.03297 | 0.077578 | -0.02773 | 0.101425 | -0.0004 | -0.0662 | -0.06258 | 0.135607 | 0.110387 | -0.11517 | -0.00964 | -0.37154 | 0.030974 | -0.0628 | 0.055896 | -0.03959 | 0.01848 | 0.06058 | 0.047806 | |
| 22 | -0.0708 | -0.02444 | 0.075415 | -0.10149 | 0.106117 | -0.00879 | -0.07141 | -0.05982 | 0.136585 | 0.111613 | -0.11603 | -0.00977 | -0.36985 | 0.026014 | -0.06221 | 0.052712 | -0.04101 | 0.031339 | 0.088012 | 0.040984 | |
| 23 | -0.05999 | -0.02516 | 0.069783 | -0.00492 | 0.102519 | -0.0138 | -0.08179 | -0.05718 | 0.130402 | 0.112386 | -0.11697 | -0.01563 | -0.36776 | 0.015382 | -0.067 | 0.04928 | -0.04159 | 0.049657 | 0.109781 | 0.037034 | |
| 24 | -0.04481 | -0.02863 | 0.062014 | 0.000132 | 0.098645 | -0.0158 | -0.0906 | -0.05462 | 0.117993 | 0.112412 | -0.11823 | -0.01996 | -0.36724 | 0.003619 | -0.07299 | 0.045728 | -0.0394 | 0.06574 | 0.123404 | 0.033951 | |
| 25 | -0.02077 | -0.02715 | 0.065596 | -0.00056 | 0.077945 | -0.01319 | -0.10081 | -0.05158 | 0.130205 | 0.112389 | -0.12017 | -0.02023 | -0.36716 | -0.00862 | -0.0808 | 0.046074 | -0.02738 | 0.082063 | 0.128988 | 0.029662 | |
| 26 | 0.01016 | -0.02506 | 0.0534 | -0.00423 | 0.024175 | -0.0038 | -0.11218 | -0.04902 | 0.087193 | 0.112009 | -0.12223 | -0.01817 | -0.36744 | -0.01917 | -0.08998 | 0.052139 | -0.01123 | 0.10054 | 0.131811 | 0.029136 | |
| 27 | 0.036032 | -0.02173 | 0.050204 | -0.00398 | -0.11693 | 0.007195 | -0.11961 | -0.04798 | 0.071409 | 0.110791 | -0.12297 | -0.01243 | -0.36836 | -0.02723 | -0.09381 | 0.061256 | -0.00078 | 0.115447 | 0.133787 | 0.024053 | |
| 28 | 0.049743 | -0.01348 | 0.046591 | -0.00201 | -0.24868 | 0.015602 | -0.12037 | -0.04802 | 0.060803 | 0.110011 | -0.12163 | -0.00598 | -0.36944 | -0.03156 | -0.09499 | 0.07456 | 0.010594 | 0.12455 | 0.132005 | 0.035603 | |
| 29 | 0.041422 | -0.02518 | 0.043214 | 0.002805 | -0.28564 | 0.022397 | -0.12124 | -0.04918 | 0.055199 | 0.109388 | -0.12198 | 0.001388 | -0.36976 | -0.02361 | -0.0946 | 0.09023 | 0.025806 | 0.128239 | 0.129201 | 0.048544 | |
| 30 | 0.020178 | -0.03777 | 0.038777 | 0.006996 | -0.29917 | 0.025193 | -0.12616 | -0.05162 | 0.050802 | 0.108011 | -0.12442 | 0.008803 | -0.36903 | -0.011 | -0.09421 | 0.100364 | 0.032171 | 0.127716 | 0.126193 | 0.005755 | |
| 31 | -0.0082 | -0.03644 | 0.031022 | 0.005186 | -0.27231 | 0.0222 | -0.12182 | -0.05478 | 0.049204 | 0.105387 | -0.12518 | 0.006378 | -0.36757 | -0.00481 | -0.09218 | 0.104024 | 0.03364 | 0.120219 | 0.121809 | -0.07616 | |
| 32 | -0.03358 | -0.04457 | 0.029796 | -0.00639 | -0.14032 | 0.025423 | -0.12101 | -0.05642 | 0.048192 | 0.102214 | -0.12692 | 0.001233 | -0.36583 | -0.00299 | -0.09123 | 0.103978 | 0.04217 | 0.109191 | 0.123004 | -0.12674 | |
| 33 | -0.05624 | -0.06164 | 0.035203 | -0.01278 | -0.01286 | 0.040185 | -0.10351 | -0.05678 | 0.043202 | 0.097983 | -0.12798 | 0.005983 | -0.36317 | -0.00461 | -0.08855 | 0.097 | 0.043802 | 0.099603 | 0.129398 | -0.15408 | |
| 34 | -0.0649 | -0.07594 | 0.038391 | -0.01764 | 0.023585 | 0.058616 | -0.06832 | -0.05682 | 0.038405 | 0.09342 | -0.12742 | 0.004985 | -0.36043 | -0.01461 | -0.08667 | 0.084004 | 0.035198 | 0.089201 | 0.135801 | -0.17627 | |
| 35 | -0.06474 | -0.08045 | 0.045626 | -0.02295 | 0.053267 | 0.080391 | -0.00237 | -0.05638 | 0.036306 | 0.088977 | -0.12579 | -0.00498 | -0.35777 | -0.02698 | -0.08792 | 0.071994 | 0.02961 | 0.079597 | 0.143002 | -0.14021 | |
| 36 | -0.08047 | -0.101 | 0.062383 | -0.01804 | 0.076463 | 0.09819 | 0.154909 | -0.05522 | 0.035607 | 0.085829 | -0.12601 | -0.0044 | -0.35483 | -0.03262 | -0.08007 | 0.059406 | 0.024583 | 0.071207 | 0.151598 | -0.0658 | |
| 37 | -0.09371 | -0.12114 | 0.075196 | -0.01838 | 0.083362 | 0.099984 | 0.335017 | -0.05479 | 0.037599 | 0.090186 | -0.12639 | -0.00864 | -0.35157 | -0.03217 | -0.03865 | 0.046994 | 0.015613 | 0.065595 | 0.155589 | -0.01766 | |
| 38 | -0.06519 | -0.12306 | 0.081406 | -0.0214 | 0.086201 | 0.084995 | 0.444464 | -0.05501 | 0.039998 | 0.094113 | -0.12782 | -0.02938 | -0.34563 | -0.02582 | -0.045488 | 0.036411 | 0.010201 | 0.062008 | 0.154209 | -0.00598 | |
| 39 | 0.107887 | -0.12534 | 0.088395 | -0.02 | 0.084423 | 0.061602 | 0.485066 | -0.05499 | 0.041003 | 0.106984 | -0.12978 | -0.046 | -0.33497 | -0.024 | -0.088361 | 0.028592 | 0.011203 | 0.059793 | 0.152191 | -0.0014 | |
| 40 | 0.33737 | -0.12426 | 0.091996 | -0.02822 | 0.08877 | 0.032986 | 0.451801 | -0.05501 | 0.041596 | 0.113817 | -0.13002 | -0.06001 | -0.32083 | -0.02038 | 0.079448 | 0.021407 | 0.019012 | 0.059411 | 0.148605 | 0.003181 | |
| 41 | 0.462688 | -0.1037 | 0.089597 | -0.04157 | 0.089828 | 0.007635 | 0.303215 | -0.055 | 0.040801 | 0.118576 | -0.12898 | -0.07879 | -0.30717 | -0.00862 | 0.075156 | 0.014795 | 0.030383 | 0.059987 | 0.143596 | 0.004028 | |
| 42 | 0.531299 | -0.07172 | 0.084202 | -0.03698 | 0.090774 | -0.00622 | 0.125087 | -0.055 | 0.038798 | 0.114808 | -0.12802 | -0.08979 | -0.29402 | 0.004827 | 0.037549 | 0.008403 | 0.040619 | 0.060213 | 0.137401 | 0.005163 | |
| 43 | 0.536339 | -0.04869 | 0.079401 | -0.03146 | 0.090019 | -0.01136 | 0.001761 | -0.0554 | 0.037204 | 0.105393 | -0.12698 | -0.09301 | -0.28479 | 0.021175 | -0.04698 | 0.004004 | 0.049776 | 0.059985 | 0.130802 | 0.002837 | |
| 44 | 0.378375 | -0.02249 | 0.075599 | -0.03473 | 0.087181 | -0.01043 | -0.07328 | -0.0564 | 0.035394 | 0.097009 | -0.12602 | -0.09339 | -0.28021 | 0.039027 | -0.09067 | 0.004401 | 0.053815 | 0.060017 | 0.124998 | 0.000564 | |

备。

工频干扰是指由于电源线的电磁干扰，心电信号中会出现 50/60Hz 的噪声（我国 50Hz 交流电的使用会产生 50Hz 的工频干扰）。这通常是由于心电图机或病人（或两者）接地不当造成的。这种噪声对于需要精确测量 P 波和 T 波的分析来说是有问题的，因为它与这些波的频率内容重叠。

基线漂移主要由呼吸和其他类型的身体缓慢移动或者电极与皮肤接触不当造成。基线漂移是一种低频噪声成分，频率通常小于 1Hz，可能导致峰值检测和分析问题。

肌电噪声是由肌肉活动产生的噪声，频率范围通常在 5~2000Hz。当在运动或剧烈身体活动期间收集心电图信号时，肌电噪声尤其成问题。其在正常心电信号上会表现为毫无规律的细小的高频波动，容易与疾病带来的小波动产生视觉混淆，从而导致心电信号诊断有误。这种噪声的去除是复杂的，因为它的频带可以与心电图的 PQRST 复合体的频带重叠。

（1）原始信号

下图为训练集前六个样本的心电图曲线，可以看到其中样本 3 和样本 4 都存在较为明显的基线漂移，且每个心电图都有细小的波动，可能是工频干扰或是肌电干扰所致。因此，需要对其进行处理，尽量消除工频干扰和基线漂移的影响。

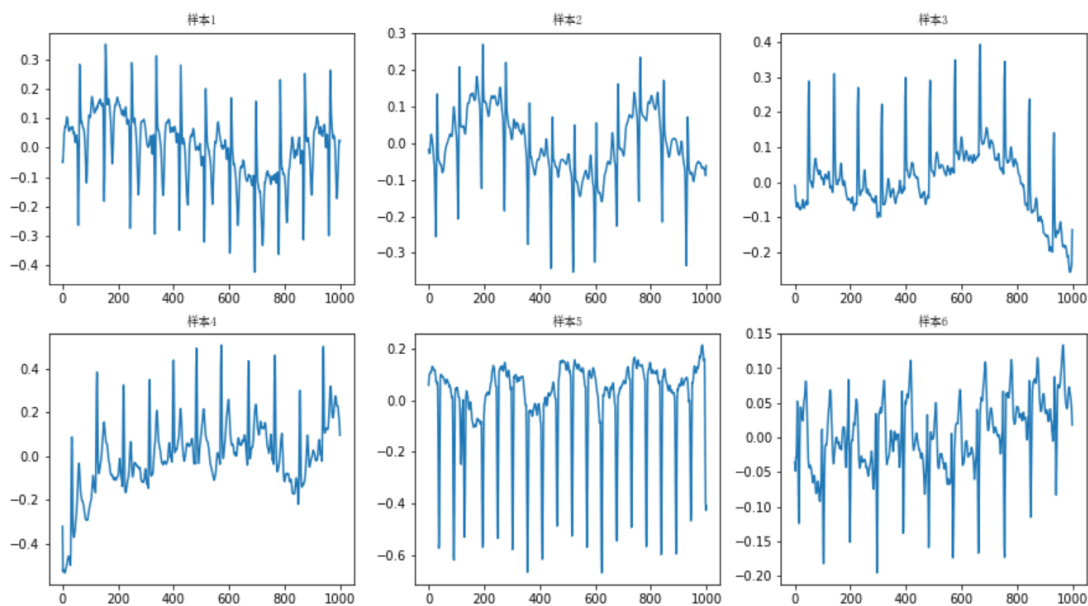


图 1 训练集心电图

心电信号发生的各种畸变是心脏以外的这些噪声干扰信号通过以乘性或加性的方式来与心电信号想混杂，这使得判断心电信号的难度增大。心电信号预处理第一考虑对原始数据进行滤波从而消除噪声影响。

（2）滤波后信号

本文采用了截止频率为 1-40Hz 的四阶巴特沃斯带通滤波器。其特点是通频带内的频率响应曲线最大限度平坦，没有起伏，而在阻频带则逐渐下降为零，在振幅的对数对角频率的波特图上，从某一边界角频率开始，振幅随着角频率的增加而逐步减少，趋向负无穷大，滤波器的阶数越高，在阻频带振幅衰减速度越快。

心电信号的频率较低，主要频率范围在 0.05-100Hz 以内，而 90%的 ECG 频谱能量集中在 0.05-40Hz 之间。由于我们的采样频率为 100Hz, 要滤除 0.05Hz 以下, 40Hz 以上频率成分，即截至频率为 0.05Hz、40Hz, 则 $W_{n1}=2*0.05/100=0.001$, $W_{n2}=2*40/100=0.8$, $W_n=[0.001, 0.8]$ 。

选择数据集中正常样本、失常样本各一个作为例子，其滤波前后对比如下所示：

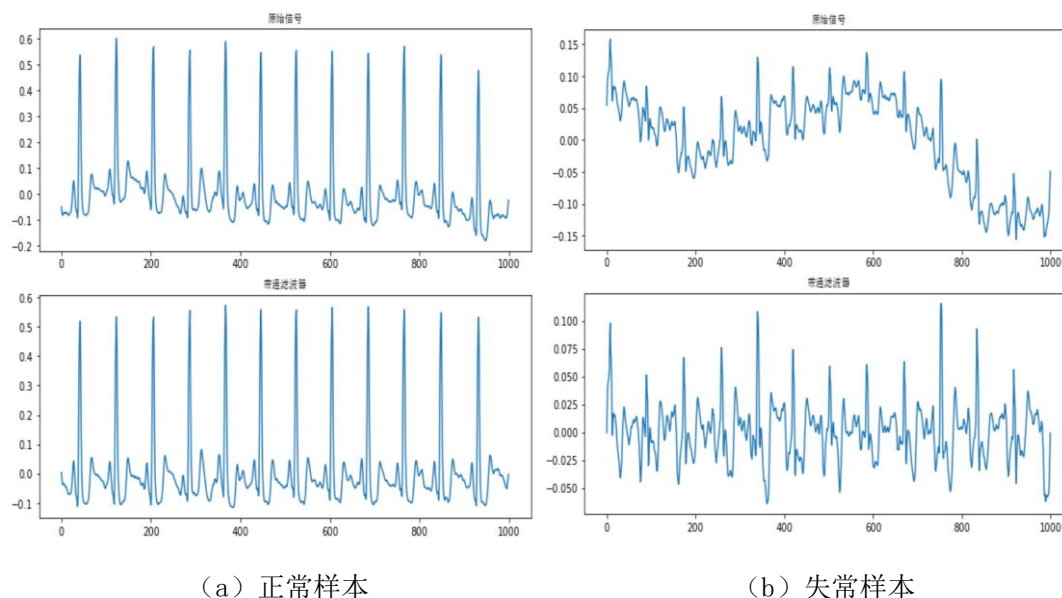


图 2 滤波前后对比

由上图可以看出，与正常样本相比，采集到的失常样本由于被采样者自身原因、机器原因及其它各种原因，存在较为明显的基线漂移情况，经过滤波处理后，该情况明显好转，有利于进一步研究。

2、R 波检测

R 波检测是心电波形分析的首要任务，只有确定这个信息之后才能对各个波段进行划分，进而提取特征。

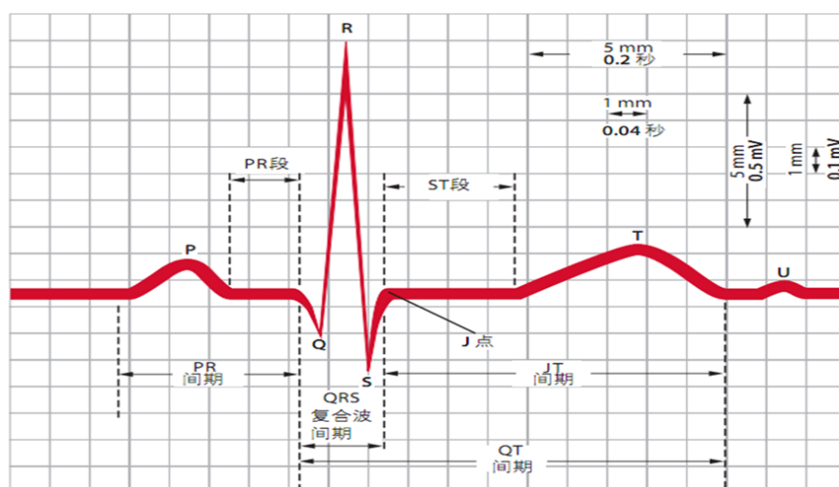


图 3 R 波

R 波的检测是通过固定差分阈值实现，基本原理是设定一个固定的峰值阈值作为限定条件，一般通过获取某一段时间内心电信号最大值和最小值，来设置捕获条件阈值，计算式如下：

$$Threshold = (\max(x) - \min(x)) \times 0.6 + \min(x)$$

仍然使用上文中选择的正常样本与失常样本，对其进行 R 波检测：

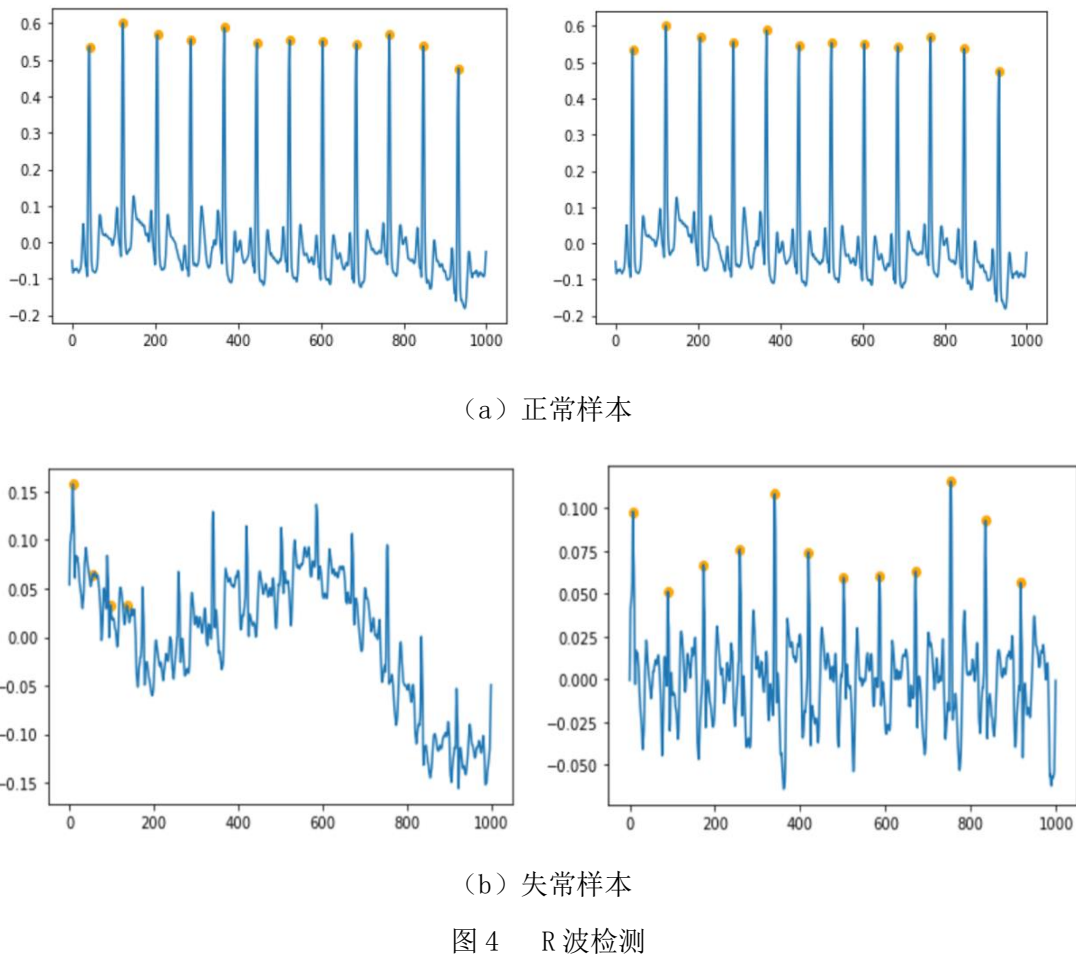


图 4 R 波检测

由上图可以看出，由于滤波对我所选择的正常样本本身影响不大，因此对滤波前后的正常样本进行 R 波检测时，结果基本一致。而对于失常样本来说，对滤波前后的 R 波检测所标注的 R 峰位置区别十分明显，对滤波后的样本进行检测时，其结果更加准确。这也可以间接说明了滤波的重要性。

3、特征提取

在 R 波检测的同时，可以截取上述两个样本中各一个波段进行对比，观察其不同复合波的差异，波段对比如下图所示：

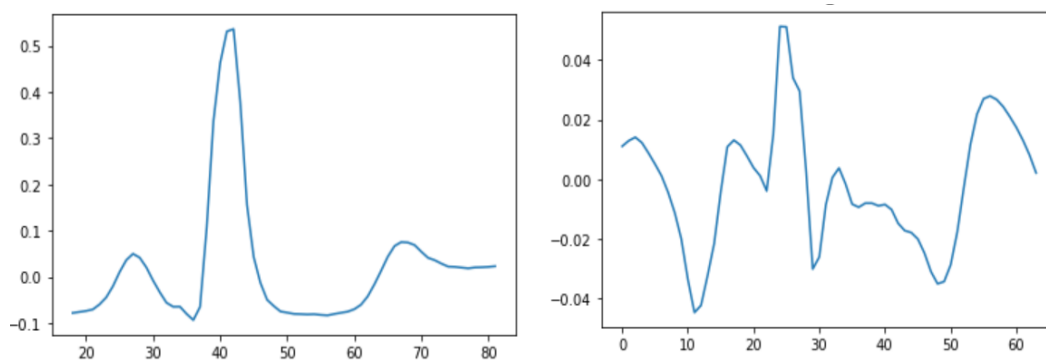


图 5 波段对比

(1) 统计学特征提取

由图 5 可以看出，失常样本的波段有着自己独特的特点，主要为 ST 段抬高、T 波高耸、Q 波加深。因此，我认为在对每一个样本进行波段划分的基础上，可以提取出每一个波段的 QRS 复合波间期和 JT 间期两个部分，对其进行探索，可提取两个波段的最小值、最大值、平均值及标准差共 8 个统计学特征。

其中 QRS 复合波片段我在 R 峰的点附近取前六个采样点与后六个采样点作为 QRS 复合波片段，JT 片段则选图 R 峰后六个采样点到后 25 个采样点作为 JT 波片段。下图为上述失常样本的 QRS 复合波间期和 JT 间期：

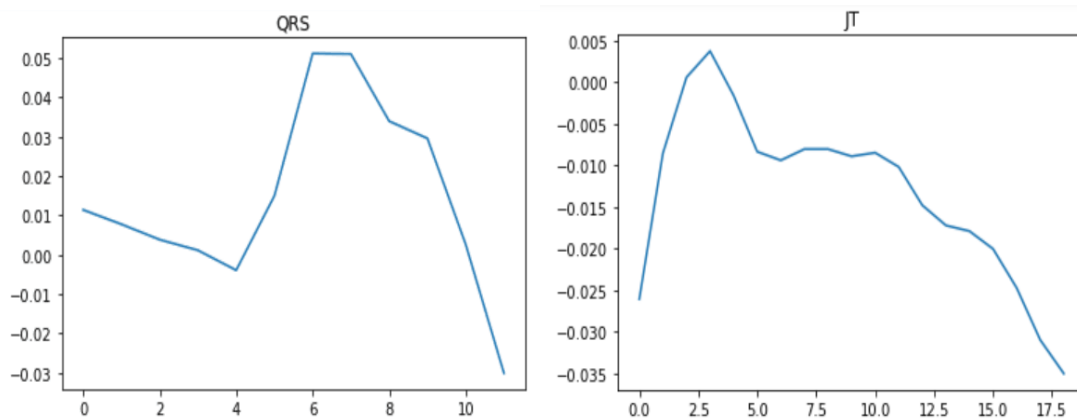


图 6 波形片段

对所有样本进行波段划分并提取 QRS 复合波间期和 JT 间期，求得两个波段的最小值、最大值、平均值及标准差，得到八个统计学特征如下：

| | MEAN_QRS | MAX_QRS | MIN_QRS | STD_QRS | MEAN_ST | MAX_ST | MIN_ST | STD_ST |
|-----|-----------|----------|-----------|----------|-----------|----------|-----------|----------|
| 0 | 0.030299 | 0.321737 | -0.344648 | 0.138103 | -0.006615 | 0.082524 | -0.191902 | 0.057435 |
| 1 | 0.002677 | 0.167135 | -0.287841 | 0.089252 | 0.007485 | 0.081532 | -0.052854 | 0.031578 |
| 2 | 0.095953 | 0.305210 | -0.079695 | 0.119660 | 0.001872 | 0.050378 | -0.035660 | 0.021710 |
| 3 | 0.109223 | 0.440166 | -0.178875 | 0.186513 | -0.011705 | 0.225889 | -0.126727 | 0.072772 |
| 4 | -0.004778 | 0.131761 | -0.614275 | 0.168288 | 0.013586 | 0.132226 | -0.633831 | 0.152244 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 175 | 0.101486 | 0.184729 | -0.012185 | 0.048789 | 0.017389 | 0.120838 | -0.077455 | 0.040971 |
| 176 | 0.031654 | 0.116505 | -0.217432 | 0.048686 | -0.009666 | 0.137519 | -0.308995 | 0.096358 |
| 177 | -0.021194 | 0.253366 | -0.410655 | 0.149612 | 0.018943 | 0.277676 | -0.359174 | 0.072749 |
| 178 | 0.071769 | 0.402933 | -0.276528 | 0.203311 | -0.032500 | 0.025320 | -0.073896 | 0.019883 |
| 179 | -0.003948 | 0.075018 | -0.151914 | 0.046903 | 0.013944 | 0.082561 | -0.035065 | 0.020744 |

180 rows × 8 columns

(2) 小波变换及特征提取

利用“Haar”小波对心电信号进行四阶小波变换，并选取一次分解得到的粗略信号对应的系数作为特征，共获得 63 个特征，如下所示：

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | ... |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----|
| 0 | -0.161441 | -0.120058 | 0.446222 | -0.364722 | 0.068091 | -0.072553 | -0.060695 | 0.467357 | -0.367100 | 0.042589 | ... |
| 1 | -0.233503 | 0.007421 | -0.154107 | 0.145596 | 0.153571 | 0.091265 | -0.219698 | -0.161247 | 0.049150 | 0.120259 | ... |
| 2 | 0.165236 | -0.035282 | 0.004191 | 0.006641 | -0.033600 | 0.041781 | -0.040362 | 0.012350 | 0.017781 | 0.035622 | ... |
| 3 | -0.453335 | -0.078892 | 0.062037 | 0.152931 | 0.223451 | -0.317446 | -0.080009 | 0.016638 | 0.119669 | 0.261404 | ... |
| 4 | 0.238280 | -0.226462 | 0.161548 | -0.201475 | 0.085875 | 0.099058 | 0.136996 | -0.256940 | 0.108023 | -0.182861 | ... |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 175 | 0.121998 | 0.109373 | -0.142246 | -0.104701 | 0.050871 | 0.097759 | 0.174970 | -0.202706 | -0.170638 | 0.077901 | ... |
| 176 | 0.121998 | 0.109373 | -0.142246 | -0.104701 | 0.050871 | 0.097759 | 0.174970 | -0.202706 | -0.170638 | 0.077901 | ... |
| 177 | 0.121998 | 0.109373 | -0.142246 | -0.104701 | 0.050871 | 0.097759 | 0.174970 | -0.202706 | -0.170638 | 0.077901 | ... |
| 178 | 0.121998 | 0.109373 | -0.142246 | -0.104701 | 0.050871 | 0.097759 | 0.174970 | -0.202706 | -0.170638 | 0.077901 | ... |
| 179 | 0.121998 | 0.109373 | -0.142246 | -0.104701 | 0.050871 | 0.097759 | 0.174970 | -0.202706 | -0.170638 | 0.077901 | ... |

200 rows × 63 columns

(3) PCA 降维

将上述所有特征加起来，一共得到 71 个特征。由于特征数量较多，因此接下来进行 PCA 进行降维处理。设置主成分个数为 20，得到其贡献率如下图所示：

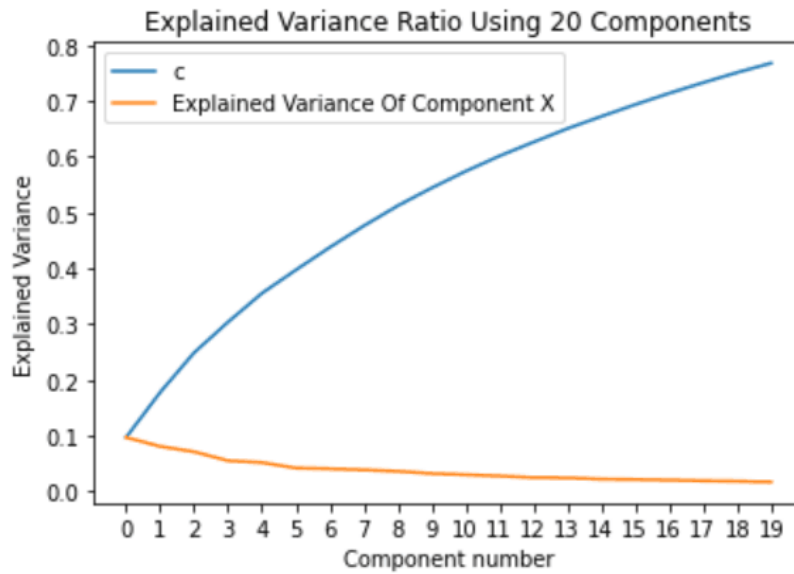


图7 主成分贡献率

由上图可以看出，所有主成分的累计贡献率接近 80%，因此，可以选择这 20 个主成分进行后续分类。最终得到的 20 个特征如下所示：

| | PC-0 | PC-1 | PC-2 | PC-3 | PC-4 | PC-5 | PC-6 | PC-7 | PC-8 | PC-9 |
|-----|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| 0 | -0.268190 | -0.018070 | 0.236031 | 0.031151 | -0.543557 | -0.026011 | 0.029334 | -0.181584 | 0.429017 | 0.461972 |
| 1 | -0.028010 | -0.274087 | 0.022056 | -0.566954 | 0.396853 | -0.115271 | 0.017256 | -0.414115 | 0.214347 | -0.012110 |
| 2 | -0.104770 | -0.118733 | -0.055673 | -0.022187 | -0.150148 | -0.196157 | -0.062401 | -0.045745 | -0.069364 | -0.152643 |
| 3 | 0.011581 | -0.144523 | 0.056685 | -0.545839 | -0.045744 | 0.257530 | -0.261786 | 0.398875 | 0.000504 | 0.526740 |
| 4 | -0.215027 | 0.172111 | -0.083609 | -0.094873 | -0.133854 | -0.078863 | 0.102501 | 0.074761 | -0.240536 | -0.305593 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 195 | -0.010345 | 0.130634 | -0.294722 | 0.017003 | -0.050033 | -0.052262 | 0.373767 | 0.106072 | -0.285392 | -0.515933 |
| 196 | -0.099444 | -0.309895 | 0.257953 | -0.178747 | 0.124602 | -0.211250 | -0.919468 | -0.264087 | -0.364399 | -0.514812 |
| 197 | 0.187733 | 0.546848 | -0.272290 | 0.128313 | 0.344808 | 0.011202 | 0.211121 | 0.016150 | 0.123875 | 0.050611 |
| 198 | 0.035106 | 0.078009 | 0.208922 | -0.293831 | -0.037314 | 0.028159 | 0.001323 | -0.266580 | 0.139260 | 0.068719 |
| 199 | 0.026090 | -0.013330 | 0.099811 | -0.060276 | 0.188611 | -0.097468 | 0.130957 | 0.082175 | -0.124092 | 0.151655 |

200 rows × 20 columns

四、心电图分类

1、SVM 算法

SVM 的核心在于找到一个超平面将两类样本准确的分开，同时保证间隔尽可能的大，这样会有更好的泛化能力。

设超平面方程为: $w^T x + b = 0$, 我们需要做的找到这样一个超平面划分两类信用状况并使得德国信贷训练集数据上的点到这个超平面的间隔距离尽可能的远。接着将原问题通过拉格朗日对偶算法转换为其对偶问题, 其主要原因是自然引入核函数从而降低求解复杂度。其对偶优化问题如下, 我们解决该优化问题即可。

$$\begin{aligned} \max L(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{s. t.} \\ &\begin{cases} \sum_{i=1}^n \alpha_i y_i = 0 \\ \alpha_i > 0, i = 1, 2, \dots, n \end{cases} \end{aligned}$$

2、评估方法

(1) K 折交叉验证

K-交叉验证是指将原始数据分成 K 组(一般是均分), 将每个子集数据分别做一次验证集, 其余的 K-1 组子集数据作为训练集, 这样会得到 K 个模型, 用这 K 个模型最终的验证集的分类准确率的平均数作为此 K-CV 下分类器的性能指标。K 一般大于等于 2, 实际操作时一般从 3 开始取, 只有在原始数据集合数据量小的时候才会尝试取 2。而 K-CV 的实验共需要建立 k 个 models, 并计算 k 次 test sets 的平均识别率。在实作上, k 要够大才能使各回合中的训练样本数够多, 一般而言 k=10 (作为一个经验参数)算是相当足够了。

这样做的好处是, 训练集的所有样本都必然会成为训练数据同时也必然有机会成为一次测试集, 可以更好的利用训练集数据。K 越大, 平均误差被视为泛化误差这个结果就越可靠。

(2) 混淆矩阵

混淆矩阵向我们展示了查准率(准确率)与查全率(召回率):

查准率(P) = $\frac{TP}{TP+FP}$, 即在被判别为正类别的样本中, 确实为正类别的比例是多少;

查全率(R) = $\frac{TP}{TP+FN}$, 即在所有正类别样本中, 被正确判别为正类别的比例是多少。

3、分类结果及对比

这里选取不同特征进行对比, 第一组为将小波变换与统计学特征一起进行 PCA 降维后的 20 个特征, 另一组为只用 8 个统计学特征。并通过 Sklearn 的 SVC 模型对心电数据集进行分类。

(1) PCA 特征

当选择 PCA 降维后的 20 个特征时, SVM 分类结果如下:

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0.0 | 0.67 | 0.73 | 0.70 | 11 |
| 1.0 | 0.62 | 0.56 | 0.59 | 9 |
| accuracy | | | 0.65 | 20 |
| macro avg | 0.65 | 0.64 | 0.64 | 20 |
| weighted avg | 0.65 | 0.65 | 0.65 | 20 |

对分类结果进行 5 折交叉验证，其 F1 值展示如下：

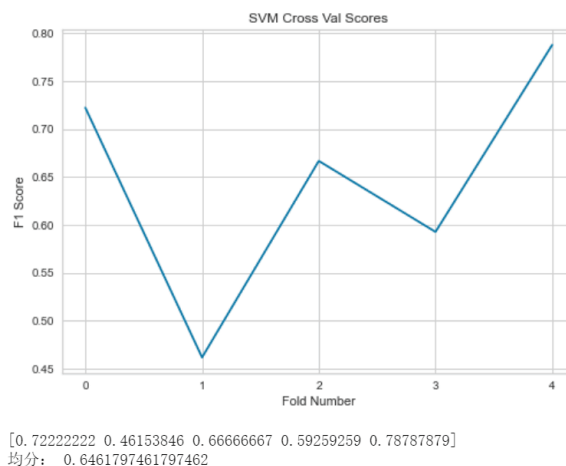


图 8 SVM 交叉验证

可以看出，模型分类的准确率仅为 65%，F1 值平均为 0.646。

接着，通过混淆矩阵来观察预测分类和实际分类情况，绘制混淆矩阵的热点图如下：

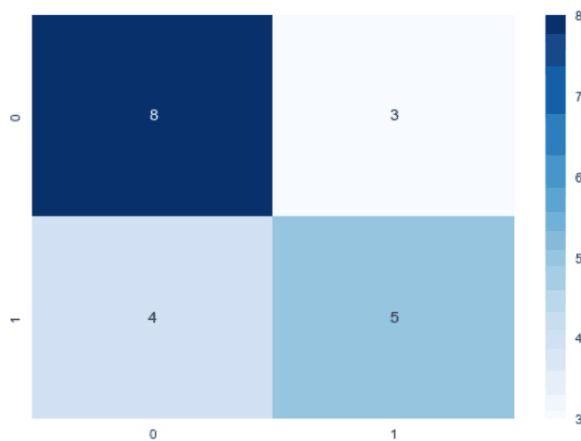


图 9 混淆矩阵

由上图可知，此模型分类效果不够理想。其中，将本应分类为正常心电（标签为 0）的 3 个样本误分为失常心电（标签为 1），同时将本应分类为失常心电（标签为 1）的 4 个样本误分为正常心电（标签为 0）。

在真正的临床应用中，将正常心电误诊为失常可能会极大影响受测者的心情及生活。同时，若将失常心电误诊为正常心电会带来更加严重的后果，极有可能耽误患者及时治疗。故

该分类效果并不理想。

(2) 统计学特征

当我们只选择 8 个统计学特征时，SVM 分类结果如下：

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| 0.0 | 0.69 | 0.82 | 0.75 | 11 |
| 1.0 | 0.71 | 0.56 | 0.63 | 9 |
| accuracy | | | 0.70 | 20 |
| macro avg | 0.70 | 0.69 | 0.69 | 20 |
| weighted avg | 0.70 | 0.70 | 0.69 | 20 |

对分类结果进行 5 折交叉验证，其 F1 值展示如下：

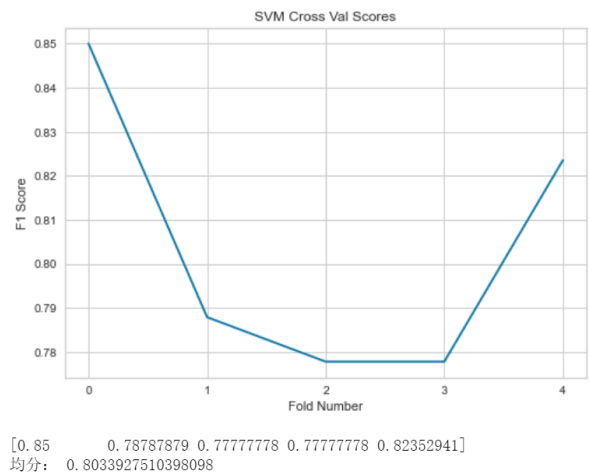


图 10 SVM 交叉验证

可以看出，模型分类的准确率为 70%，F1 值平均为 0.803。与第一组特征相比，准确率和 F1 值都有所提高。

接着，通过混淆矩阵来观察预测分类和实际分类情况，绘制混淆矩阵的热点图如下：

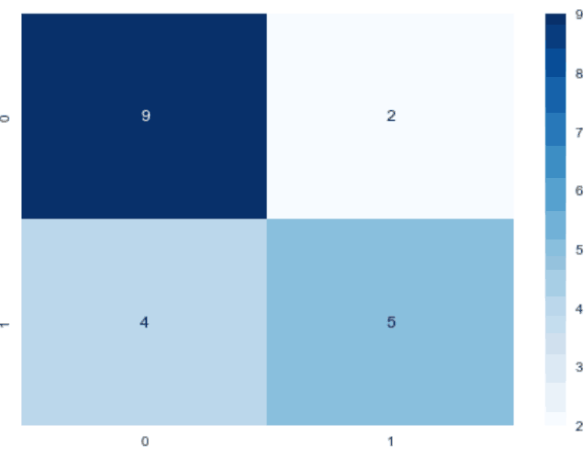


图 11 混淆矩阵

由上图可知，此模型相较于第一组而言相对理想。其中，将本应分类为正常心电（标签为 0）的 2 个样本误分为失常心电（标签为 1），同时将本应分类为失常心电（标签为 1）的 4 个样本误分为正常心电（标签为 0）。可以明显看出正常心电划分为失常心电的误分类明显减少，但失常心电误分为正常心电这更严重的错误并没有被减少，说明该分类仍有弊端。

五、实践总结

本文以心电数据集为实验数据，采用 SVM 对数据集进行分类。首先，对心电信号数据进行滤波处理，去除工频干扰和基线漂移。其次，对滤波后的心电信号进行 R 波检测并分段，根据不同复合波的特点，分别对 QRS 复合波间期和 JT 间期进行统计学特征提取。之后，再利用小波变换提取特征，将所有特征汇总后利用 PCA 进行降维处理，以此进行 SVM 分类。最后，对比了使用 PCA 降维后的 20 个特征和 8 个统计学特征这两组不同特征进行分类时的 SVM 分类结果。发现利用 8 个统计学特征准确率高于 20 个特征的准确率，这可能是因为我小波变换提取特征时，没有选择到较为合适的分解阶数对应的系数，因此 8 个统计学特征的分类准确率更高。但在实际生活诊断中，将失常心电误分类为正常心电的后果十分严重。上述两种分裂结果对于这个误分类问题的效果相似，由此可见两者结果均需要改进。

通过本次实验，我对于人的心电信号有了初步的认识和理解；其次，我学习到了如何对心电信号进行一系列处理，这个过程不但加深了我对于小波变换理论的理解，对波段划分的认识，同时极大地锻炼了我的动手实践能力；此外，运用 SVM 对数据进行分类，再一次巩固了对该算法的掌握与运用。

但是这次对于心电信号的分类中，利用两组特征的分类结果均甚不理想，其主要问题在于我对于小波变换提取特征这里还存在改进空间；其次，对于波段的划分也许不够十分准确，利用人工肉眼进行，缺乏相对应的理论支撑；最后，在分类方法上我也需要变换方法进行尝试。在之后的实验中，可以在这些方面改进实验方法，以得到更好的结果。